



UNIVERSIDAD ANDINA
NESTOR CACERES VELASQUEZ
FACULTAD DE INGENIERÍA DE SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



**APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR
LA EFICIENCIA EN LA PESCA ARTESANAL DE LA
EMPRESA PUERTO LAGO PUNO 2024**

TESIS PRESENTADA POR:
Bach. ALBERT LUIGHI FIGUEROA CARI

PARA OPTAR EL TÍTULO PROFESIONAL DE:
INGENIERO DE SISTEMAS

JULIACA – PERÚ
2026



UNIVERSIDAD ANDINA

NÉSTOR CÁCERES VELÁSQUEZ

FACULTAD DE INGENIERÍA DE SISTEMAS

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

**APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR
LA EFICIENCIA EN LA PESCA ARTESANAL DE LA
EMPRESA PUERTO LAGO PUNO 2024**

TESIS PRESENTADA POR:

Bach. ALBERT LUIGHI FIGUEROA CARI

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS

APROBADA POR EL JURADO REVISOR:

PRESIDENTE

:


M. Sc. JUAN CARLOS HERRERA MIRANDA

PRIMER MIEMBRO

:


MSc. JESUS ESTEBAN CASTILLO MACHACA

SEGUNDO MIEMBRO

:


Ing. ADWAR RANULFO SANCHEZ CARREON

ASESOR DE TESIS

:


Dr. PAUL MAMANI TISNADO

LÍNEA DE INVESTIGACIÓN:

CIENCIA DE LOS ORDENADORES – P24

**RESOLUCIÓN DECANAL N° 088-2026-D-FIS-UANCV**

Juliaca, 28 de enero de 2026

VISTOS; el Expediente N° 2026-CU-1096, presentado por el señor (a) **ALBERT LUIGHI FIGUEROA CARI**, egresado de la Escuela Profesional de INGENIERIA DE SISTEMAS, de la Facultad de Ingeniería de Sistemas, quien solicita **NOMINACIÓN DE JURADOS Y PROGRAMACIÓN DE FECHA Y HORA DE SUSTENTACIÓN**.

CONSIDERANDO:

Que, el (la) Bachiller **ALBERT LUIGHI FIGUEROA CARI**, quien solicita **NOMINACIÓN DE JURADOS Y PROGRAMACIÓN DE FECHA Y HORA DE SUSTENTACIÓN** de la tesis titulada: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, la misma que pertenece a la línea de investigación, CIENCIA DE LOS ORDENADORES, para optar el título profesional de INGENIERO DE SISTEMAS.

Que, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación conducente a Grados y Títulos mediante resolución 0294-2023-UANCV-CU-R y en concordancia con el dictamen de similitud.

De conformidad al Reglamento Interno de Trabajos de Investigación conducente a grados y títulos, aprobado con resolución N° 0294-2023-UANCV-CU-R y en mérito al art. 24, Art. 28 del reglamento, con fines de obtención de Grados Académicos y Títulos Profesionales y en uso a las atribuciones que le concede la ley Universitaria N° 30220, Ley de creación de la UANCV N° 23738 y modificatoria N° 24661, y el estatuto de la UANCV, el Decano y Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas.

RESUELVE:

ARTÍCULO PRIMERO. - APROBAR la **NOMINACION DE JURADOS**, integrado por los siguientes docentes:

- **Presidente : Dr. JUAN CARLOS HERRERA MIRANDA**
- **1er miembro : M.Sc. JESUS ESTEBAN CASTILLO MACHACA**
- **2do miembro : Ing. ADWAR RANULFO SANCHEZ CARREON**

ARTÍCULO SEGUNDO.- RECONOCER como asesor de la investigación (tesis) de la Facultad de Ingeniería de Sistemas, al (a la) docente, **Dr. PAUL MAMANI TISNADO**

ARTÍCULO TERCERO.- APROBAR la **FECHA Y HORA DE SUSTENTACION DE LA TESIS** del (la) Bachiller: **ALBERT LUIGHI FIGUEROA CARI**, del informe final de la investigación (tesis) titulado: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, para optar el título profesional de INGENIERO DE SISTEMAS, de acuerdo al siguiente detalle:

- **FECHA : jueves 29, de enero de 2026**
- **HORA : 12:30:00 PM**
- **LUGAR : Sala de Docentes-FIS**

ARTÍCULO CUARTO. - DISPONER que, la Unidad de Investigación, Responsables del Comité de Investigación de la Facultad de Ingeniería de Sistemas y el Director de la Escuela profesional, INGENIERIA DE SISTEMAS, quedan encargados del cumplimiento de la presente Resolución.

Regístrese, Comuníquese y Archívese.

UNIVERSIDAD ANDINA
"NÉSTOR CÁCERES VELÁSQUEZ"Dr. Juan Carlos Herrera Miranda
DECANO



RESOLUCIÓN DECANAL N° 149-2025-D-FIS-UANCV

Juliaca, 30 de diciembre de 2025

VISTOS; el Expediente N° 2025-CU-11984, presentado por el (a) señor (a) **ALBERT LUIGHI FIGUEROA CARI**, quien solicita **REVISION DEL INFORME FINAL DE LA INVESTIGACIÓN** (borrador de tesis), el proveído N° 071 -2025-UI-FIS-UANCV/J, del Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas.

CONSIDERANDO:

Que, el (la) señor (a) **ALBERT LUIGHI FIGUEROA CARI**, ha presentado su informe final de la investigación (borrador de tesis), titulado: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, para optar el título profesional de **INGENIERO DE SISTEMAS**.

Que, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación conducente a Grados y Títulos con fines de obtención de Grados Académicos y Títulos, el Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas emitió opinión favorable al Expediente 2025-CU-11984, del informe final (borrador de tesis);, titulado: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, correspondiente a la línea de investigación, **CIENCIA DE LOS ORDENADORES**.

Que al haberse cumplido con los requisitos exigidos por el reglamento interno de trabajos de investigación conducentes a grados y títulos mediante resolución N° 0294-2023-UANCV-CU-R, y estando a la opinión favorable de la Unidad de Investigación respecto al informe final de la investigación (borrador de tesis).

Estando, con la opinión favorable de la propuesta de investigación del Comité de investigación de la Facultad de Ingeniería de Sistemas y en concordancia con el reglamento interno de trabajos de investigación conducente a grados y títulos, aprobado con resolución N° 0294-2023-UANCV-CU-R y en mérito al art. 25 del reglamento, con fines de obtención de obtención de Grados Académicos y Títulos Profesionales y en uso a las atribuciones que le concede la ley Universitaria N° 30220, Ley de creación de la UANCV N° 23738 y modificatoria N° P 24661, y el estatuto de la UANCV, el Decano y Director de la Facultad de Ingeniería de Sistemas.

RESUELVE:

ARTÍCULO PRIMERO.- APROBAR, el **INFORME FINAL DE LA INVESTIGACION** (BORRADOR DE TESIS), para la **REVISION DE SIMILITUD TURNITIN**, presentado por el (la) señor (a): **ALBERT LUIGHI FIGUEROA CARI**, para optar el Título Profesional de **INGENIERO DE SISTEMAS** con el tema titulado: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, correspondiente a la línea de investigación, **CIENCIA DE LOS ORDENADORES**, en virtud a los considerandos expuestos.

ARTÍCULO SEGUNDO. - RATIFICAR como **ASESOR DE INVESTIGACION** al (a la) **Dr. Paul Mamani Tisnado**.

ARTÍCULO TERCERO. - DISPONER que, la Unidad de Investigación, Responsables del Comité de Investigación de la Facultad de Ingeniería de Sistemas y el Director de la Escuela profesional, **INGENIERIA DE SISTEMAS**, quedan encargados del cumplimiento de la presente Resolución.

Regístrese, Comuníquese y Archívese.



UNIVERSIDAD ANDINA
"NÉSTOR CÁCERES VELÁSQUEZ"

Dr. Juan Carlos Herrera Miranda
DECANO

**RESOLUCIÓN DECANAL N° 092-2025-D-FIS-UANCV**

Juliaca, 29 de diciembre de 2025

VISTOS; el Expediente N° 2025-CU-15272, presentado por el señor (a) **ALBERT LUIGHI FIGUEROA CARI** solicitando **CAMBIO DE ASESOR DE INVESTIGACIÓN**, el PROVEIDO N° 050-2024-UI-FIS-UANCV/J del Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas, y la RESOLUCION N° 065-2025-UI-P-D-FIS-UANCV-J, aprobación de la PROPUESTA DE INVESTIGACION, para optar el título profesional de **INGENIERO DE SISTEMAS**

CONSIDERANDO:

Que, el (la) señor (a) **ALBERT LUIGHI FIGUEROA CARI**, ha presentado su cambio de asesor de tesis del tema de investigación titulado: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, para optar el Título Profesional de **INGENIERO DE SISTEMAS**.

Que, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación conducente a Grados y Títulos con fines de obtención de Grados Académicos y Títulos, el Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas emitió opinión favorable de la propuesta de investigación; aprobando la PROPUESTA DE INVESTIGACION titulada: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**.

Que el Director de la Unidad de Investigación de la FIS ha tomado conocimiento que el (la) asesor (ra) Mtro. **ABELARDO LEON MIRANDA**, no tiene vínculo laboral en la Facultad de Ingeniería de Sistemas y la Universidad Andina Néstor Cáceres Velásquez; existiendo la RESOLUCION N°065-2025 UI-P-D-FIS-UANCV-J, aprobación de la PROPUESTA DE INVESTIGACION.

Que es requisito indispensable contar con un asesor docente ordinario y/o contratado de la facultad de Ingeniería de Sistemas, con mínimo de cinco años de docencia, grado de doctor o magister y experiencia en la línea a investigar, o deberá estar acreditado por resolución N° 0989-2022-UANCV-CU-R, quien asumirá como asesor de la propuesta de investigación según el área de grado.

Estando, con la opinión favorable del Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas y en concordancia con el reglamento interno de trabajos de investigación conducente a grados y títulos, aprobado con resolución N° 0294-2023-UANCV-CU-R y en mérito al art. 25 del reglamento, con fines de obtención de obtención de Grados Académicos y Títulos Profesionales y en uso a las atribuciones que le concede la ley Universitaria N° 30220, Ley de creación de la UANCV N° 23738 y modificatoria N° 24661, y el estatuto de la UANCV, el Decano y Director de la Facultad de Ingeniería de Sistemas.

RESUELVE:

ARTÍCULO PRIMERO.- APROBAR, el CAMBIO DE ASESOR DE INVESTIGACIÓN, presentada por el (la) señor (a): **ALBERT LUIGHI FIGUEROA CARI**, para optar el Título Profesional de **INGENIERO DE SISTEMAS** con el tema titulado: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, se le asigna como:

ASESOR: Dr. PAUL MAMANI TISNADO

ARTÍCULO SEGUNDO. - RECONOCER como **ASESOR DE INVESTIGACION** al (a) la) docente **Dr. PAUL MAMANI TISNADO**.

ARTÍCULO TERCERO. - DISPONER que, la Unidad de Investigación, Responsables del Comité de Investigación de la Facultad de Ingeniería de Sistemas y el Director de la Escuela profesional, **INGENIERIA DE SISTEMAS**, quedan encargados del cumplimiento de la presente Resolución.

Regístrese, Comuníquese y Archívese.



UNIVERSIDAD ANDINA
"NÉSTOR CÁCERES VELÁSQUEZ"
Dr. Juan Carlos Herrera Miranda



RESOLUCIÓN N° 065-2025-UI.P-D-FIS-UANCV-J

Juliaca, 20 de junio de 2025

VISTOS:

El Expediente: 2025-006857 de fecha 13 de junio de 2025, del (la) Bach. **ALBERT LUIGHI FIGUEROA CARI**; con el cual solicita Revisión de la Propuesta de Investigación y el Anexo (02 o 03) "Ficha de Opinión de la Propuesta de Investigación" que fue revisada por el Comité de Investigación de la Facultad de Ingeniería de Sistemas, Escuela Profesional de INGENIERÍA DE SISTEMAS.

CONSIDERANDO:

Que, las Unidades de Investigación son unidades académicas que agrupan a docentes y estudiantes de diversas disciplinas, en razón del desarrollo de investigación científica, tecnológica y humanista de acuerdo al Estatuto Universitario Modificado 2020 de nuestra primera Casa Superior de Estudios.

Que, el (la) Bach. **ALBERT LUIGHI FIGUEROA CARI**, solicito la revisión y aprobación de la Propuesta de Investigación de la tesis titulada: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**; conducente para optar el Título Profesional de INGENIERO DE SISTEMAS.

Que, al haberse cumplido con los requisitos exigidos por el Reglamento Interno de Trabajo de Investigación Conducente a Grados y Títulos plasmado en la Resolución N° 0294-2023-UANCV-CU-R.

Que, el Comité de Investigación ha emitido opinión favorable a la propuesta de investigación.

Que, el Director de la Unidad de Investigación de la Facultad de Ingeniería de Sistemas, Escuela Profesional de INGENIERÍA DE SISTEMAS, ratificó la propuesta del Asesor Mtro. **ABELARDO LEON MIRANDA**, quien debe estar acreditado y facultado para orientar y ayudar al asesorado en el proceso de elaboración del trabajo de investigación (Tesis).

Estando, la opinión favorable del comité de Investigación, en concordancia con el Reglamento Interno de Trabajo de Investigación Conducente a Grados y Títulos, Resolución N° 0294-2023-UANCV-CU-R, de conformidad a lo que establece la Ley Universitaria N° 30220, Ley de Creación de la UANCV N° 23738 y Modificatoria N° 24661 y el Estatuto de la UANCV, que confiere facultades al Decano de la Facultad de Ingeniería de Sistemas.

SE RESUELVE:

ARTÍCULO PRIMERO. - APROBAR Y AUTORIZAR LA EJECUCIÓN DE LA PROPUESTA DE INVESTIGACIÓN, titulada: **APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024**, presentado por el (la) Bach. **ALBERT LUIGHI FIGUEROA CARI**, para optar el Título Profesional de INGENIERO DE SISTEMAS, en virtud de los considerandos expuestos.

ARTÍCULO SEGUNDO. - RECONOCER, como **ASESOR** al Mtro. **ABELARDO LEON MIRANDA**.

ARTÍCULO TERCERO. - DISPONER que la facultad, secretarías académicas y administrativas, quedan encargados del cumplimiento de la presente resolución.

Regístrese, comuníquese y archívese.



UNIVERSIDAD ANDINA
"NÉSTOR CÁCERES VELÁSQUEZ"

M.Sc. Juan Carlos Herrera Miranda
DECANO

C.c
Arch 2025
JCHM/ v1.2
Distribución: Asesor de Tesis, Interesado

Ciudad Universitaria Urbanización Taparachi Km. 1.5 Salda Puno - Juliaca



ALBERT LUIGHI FIGUEROA CARI

APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA P...

 Trabajos de tesis para ingreso a repositorio

Detalles del documento

Identificador de la entrega

trn:oid:::13016:591409169

Fecha de entrega

15 may 2026, 23:36 GMT-5

Fecha de descarga

16 may 2026, 0:03 GMT-5

Nombre del archivo

T036_70238459_T.docx

Tamaño del archivo

12.1 MB

118 páginas

21.256 palabras

123.590 caracteres



13% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Coincidencias menores (menos de 10 palabras)

Fuentes principales

- 9%  Fuentes de Internet
- 2%  Publicaciones
- 12%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.



Metadatos complementarios



Título de la Tesis	
APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024	
Datos de autor	
Nombres y apellidos	ALBERT LUIGHI FIGUEROA CARI
Tipo de documento de identidad	DNI
Numero de documento de identidad	70238459
URL de ORCID	https://orcid.org/0009-0004-8297-0285
Datos de asesor	
Nombres y apellidos	PAUL MAMANI TISNADO
Tipo de documento de identidad	DNI
Numero de documento de identidad	01314987
URL de ORCID	https://orcid.org/0000-0002-0287-7143
Datos de jurado	
Presidente del jurado	
Nombres y apellidos	JUAN CARLOS HERRERA MIRANDA
Tipo de documento de identidad	DNI
Numero de documento de identidad	29606930
Miembro del jurado 1	
Nombres y apellidos	JESUS ESTEBAN CASTILLO MACHACA
Tipo de documento de identidad	DNI
Numero de documento de identidad	01323821
Miembro del jurado 2	
Nombres y apellidos	ADWAR RANULFO SANCHEZ CARREON
Tipo de documento de identidad	DNI
Numero de documento de identidad	02064066



Datos de investigación	
Línea de investigación	Ciencia de los ordenadores – P24
Grupo de investigación	No aplica.
Agencia de financiamiento	Sin financiamiento.
Ubicación geográfica de la investigación	País: Perú Departamento: Puno Provincia: Puno Distrito: Puno EMPRESA: PUERTO LAGO Coordenadas: Latitud: -15.839998 Longitud: -70.022043 URL Maps: https://maps.app.goo.gl/FmUxrFu6kkbjbVXA 
Año o rango de años en que se realizó la investigación	Julio 2025 – Enero 2026
URL de disciplinas OCDE https://concytec-pe.github.io/Peru-CRIS/vocabularios/ocde_ford.html	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.01.04 Ingeniería de procesos https://purl.org/pe-repo/ocde/ford#2.04.02



UNIVERSIDAD ANDINA
"NÉSTOR CACERES VELÁSQUEZ"

Dr. Juan Carlos Herrera Miranda
DIRECTOR (e)
Unidad de Investigación FIS

DECLARACIÓN DE AUTENTICIDAD Y RESPONSABILIDAD

Yo ALBERT LUIGHI FIGUEROA CARI, identificado con DNI
Nro. 70238459, en mi condición de egresado de:

- ☐ Escuela Profesional
☒ Programa de Segunda Especialidad,
☐ Programa de Maestría o Doctorado

INGENIERÍA DE SISTEMAS

informo que he elaborado el/la ☐ Tesis o ☐ Trabajo de Investigación, ☒ Trabajo Académico
denominada:

APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA
EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024

Asesorado por: Dr. PAUL MAMANI TISNADO

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

El incumplimiento de lo declarado da lugar a responsabilidad del declarante, en consecuencia; a través del presente documento asumo frente a terceros, la Universidad Andina Néstor Cáceres Velásquez y/o la Administración Pública toda responsabilidad que pueda derivarse por el trabajo final presentado. Lo señalado incluye responsabilidad pecuniaria incluido el pago de multas u otros por los daños y perjuicios que se ocasionen.

Juliaca 06 de ABRIL del 2026



Firma del Asesor
(obligatoria)



Firma del Estudiante
(obligatoria)



Huella



DEDICATORIA

A mis padres, Alberto y Elsa, quienes nunca soltaron mi mano en los momentos más difíciles de mi vida; su amor incondicional ha sido mi mayor fortaleza. A mi hermana Zindy, por ser siempre mi ejemplo a seguir y mi guía constante. Y a mis amigos, Antonieta y Daniel, piezas fundamentales en mi existencia, con quienes encontré el camino y un cariño mutuo que atesoro profundamente.



AGRADECIMIENTO

En primer lugar, agradezco a la Universidad por brindarme los conocimientos y el espacio académico para mi formación profesional.

Mi profundo agradecimiento a la empresa Puerto Lago Puno por permitirme el acceso a sus datos y confiar en la aplicación de nuevas tecnologías para la pesca artesanal.

Asimismo, agradezco a mis asesores y docentes, quienes con su guía y rigor científico orientaron el desarrollo.

Finalmente, agradezco a todas las personas que, de manera directa o indirecta, aportaron para culminar con éxito este proyecto de investigación.

.



ÍNDICE

DEDICATORIA	i
AGRADECIMIENTO	ii
ÍNDICE	iii
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS.....	vii
RESUMEN.....	viii
ABSTRACT	ix
INTRODUCCIÓN.....	x

CAPÍTULO I

ASPECTOS GENERALES

1.1. Descripción del problema	13
1.1.1 A nivel Internacional	13
1.1.2 A nivel nacional	14
1.1.3 A nivel local	14
1.2. Formulación del problema	15
1.2.1 Problema Principal	15
1.2.2 Problemas específicos	15
1.3. Justificación de la investigación	16
1.3.1 Justificación Teórica	16
1.3.2 Justificación practico	16
1.3.3 Justificación Metodológica.....	17
1.4. Objetivos.....	18
1.4.1 Objetivo general	18



1.4.2	Objetivos específicos	18
1.5.	Importancia y alcance	19
1.6.	Limitaciones y delimitaciones.....	20
1.7.	Hipótesis	22
1.7.1	Hipótesis General.....	22
1.7.2	Hipótesis Específicas	22
1.8.	Variables e indicadores.....	23
1.8.1	Conceptualización de las variables Variable 1: Algoritmo Modelo Predictivo	23

CAPÍTULO II

FUNDAMENTOS TEÓRICOS

2.1.	Antecedentes	24
2.1.1	Internacionales	24
2.1.2	Nacionales.....	27
2.1.3	Locales	30
2.2.	Marco epistemológico	33
2.3.	Estado del arte.....	34
2.4.	Bases teóricas	37
2.5.	Marco conceptual.....	54

CAPÍTULO III

METODOLOGÍA Y RESULTADOS

3.1.	Enfoque de la investigación	59
3.2.	Métodos aplicados de investigación	59
3.3.	Tipo de investigación	60



3.4. Nivel.....	61
3.5. Diseño.....	61
3.6. Ámbito de investigación	62
3.7. Población y muestra	65
3.7.1 Población.....	65
3.7.2 Muestra	65
3.8. Métodos y técnicas de recogida de información	66
3.8.1 Técnicas de investigación	66
3.8.2 Instrumento de investigación.....	68
3.8.3 Criterios de Inclusión.....	69
3.8.4 Criterios de Exclusión.....	70
3.9. Validez y confiabilidad del instrumento de investigación	71
3.9.1 Validación de los instrumentos	71
3.9.2 Confiabilidad de los instrumentos.....	72

CAPÍTULO IV

RESULTADOS

4.1. Presentación, análisis e interpretación de datos.....	74
4.2. Discusión de resultados.....	103
CONCLUSIONES	105
RECOMENDACIONES.....	107
REFERENCIAS BIBLIOGRÁFICAS.....	110
APÉNDICES.....	112



ÍNDICE DE TABLAS

Tabla 1 ¿El registro histórico contiene información sobre la fecha de la faena de pesca?	81
Tabla 2 ¿El registro consigna el volumen de captura?	82
Tabla 3 ¿El registro contiene información meteorológica (viento, precipitación)?	83
Tabla 4 ¿El modelo predictivo alcanza un coeficiente de determinación R^2 mayor a 0.75?	84
Tabla 5 ¿El Error Cuadrático Medio (MSE) obtenido es aceptable para la predicción?	85
Tabla 6 ¿El uso del modelo predictivo mejora la estimación del volumen de captura diaria?.....	86
Tabla 7 ¿La predicción del modelo permite una mejor planificación de las faenas de pesca?.....	87
Tabla 8 ¿La implementación del modelo reduce la incertidumbre en la toma de decisiones operativas?	88
Tabla 9 En comparación con el método tradicional, la eficiencia operativa de la pesca artesanal	102



ÍNDICE DE FIGURAS

Figura 1 Estructura del DataFrame	75
Figura 2 Selección de variables.....	75
Figura 3 evaluación de modelos.....	76
Figura 4 comparación de modelos	77
Figura 5 valores nulos por columna.....	78
Figura 6 código del histograma	79
Figura 7 Matriz de confusión o correlacion.....	79
Figura 8 Disponibilidad de información sobre la fecha de la faena.....	81
Figura 9 registro del volumen de captura en la base de datos histórica.....	82
Figura 10 Disponibilidad de información meteorológica	83
Figura 11 Rendimiento del modelo predictivo.....	84
Figura 12 Aceptabilidad del Error Cuadrático Medio	85
Figura 13 Impacto del modelo predictivo	86
Figura 14 Contribución del modelo predictivo a la planificación de las faenas de pesca.....	87
Figura 15 Reducción de la incertidumbre en la toma de decisiones operativas ...	88
Figura 16 eficiencia operativa de la pesca artesanal frente al método tradicional	102



RESUMEN

Esta investigación aplicada y de diseño tecnológico tiene como objetivo principal diseñar e implementar un modelo predictivo basado en algoritmos de machine learning para optimizar la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno en el Lago Titicaca. Para ello, se sigue un enfoque cuantitativo con un diseño no experimental de modelado predictivo, trabajando sobre una población de 7,560 registros históricos de captura, meteorología y limnología del periodo 2018-2023. La metodología se fundamenta en el ciclo de ciencia de datos, empleando técnicas de experimentación computacional con el lenguaje Python y librerías especializadas (Pandas, Scikit-learn) para el preprocesamiento, la ingeniería de características, el entrenamiento y la validación comparativa de algoritmos como Random Forest y XGBoost. El modelo final se evaluará con métricas de regresión (RMSE, R^2) en un conjunto de prueba aislado. Los resultados esperados incluyen la identificación de las variables ambientales más determinantes, un modelo validado con precisión satisfactoria ($R^2 > 0.75$) para estimar la biomasa o la Captura por Unidad de Esfuerzo (CPUE), y una evaluación de su impacto potencial en indicadores de eficiencia como la reducción del esfuerzo infructuoso y la mejora de la productividad. La investigación contribuye a cerrar la brecha tecnológica en el sector pesquero artesanal continental, ofreciendo una herramienta de soporte a la decisión basada en datos para una planificación más racional y sostenible.

Palabras clave: Machine learning, pesca artesanal, Lago Titicaca, modelo predictivo, eficiencia operativa, ciencia de datos, captura por unidad de esfuerzo (CPUE), Python, Random Forest, sostenibilidad pesquera.



ABSTRACT

This applied and technological design research aims to design and implement a predictive model based on machine learning algorithms to optimize the operational efficiency of the artisanal fishing operations of Empresa Puerto Lago Puno in Lake Titicaca. A quantitative approach with a non-experimental predictive modelling design is followed, working on a population of 7,560 historical records of catch, meteorology, and limnology from the 2018-2023 period. The methodology is based on the data science lifecycle, employing computational experimentation techniques with the Python language and specialized libraries (Pandas, Scikit-learn) for preprocessing, feature engineering, training, and comparative validation of algorithms such as Random Forest and XGBoost. The final model will be evaluated with regression metrics (RMSE, R^2) on an isolated test set. Expected results include identifying the most determinant environmental variables, a validated model with satisfactory accuracy ($R^2 > 0.75$) for estimating biomass or Catch per Unit Effort (CPUE), and an assessment of its potential impact on efficiency indicators such as reducing futile effort and improving productivity. The research contributes to closing the technological gap in the continental artisanal fishing sector by providing a data-driven decision-support tool for more rational and sustainable planning.

Keywords: Machine learning, artisanal fishing, Lake Titicaca, predictive model, operational efficiency, data science, catch per unit effort (CPUE), Python, Random Forest, fisheries sustainability.



INTRODUCCIÓN

La pesca artesanal es fundamental para muchas comunidades, ya que proporciona seguridad alimentaria, empleo y preservación cultural. En Perú, la pesca en el lago Titicaca es el sustento de la comunidad andina de Puno, pero enfrenta dificultades para mantenerse, principalmente debido a la alta incertidumbre operativa. Los planes de pesca dependen de las habilidades y la experiencia de los pescadores, las cuales son reactivas y susceptibles a los cambios ambientales. Esta dependencia genera ineficiencia económica, costos improductivos y una asignación inadecuada de recursos, lo que ejerce presión sobre la rentabilidad y la conservación, sin que esto se ajuste a la disponibilidad potencial de recursos.

Frente a este dilema aparece la Inteligencia Artificial (IA), especialmente el Aprendizaje Automático, como una tecnología disruptiva capaz de revolucionar los antiguos sectores de producción. Estos algoritmos pueden analizar enormes cantidades de datos históricos, detectar relaciones complejas y realizar predicciones precisas en áreas como la agricultura de precisión, la logística y la pesca oceánica industrial.

Existe una gran brecha en la investigación y el uso de la investigación y el uso artesanal continental, y la utilización de estas herramientas aún se encuentra en sus primeras etapas. de la pesca artesanal continental, y el uso de estas herramientas aún se encuentra en sus primeras etapas. ideas sobre el lago Titicaca es un desafío científico y técnico único: ¿cómo utilizar los últimos métodos en una situación en la que no hay muchos datos, hay factores ambientales específicos y las operaciones son a pequeña escala? Pasos ¿Qué ¿Podemos? tomar para encontrar una solución nueva para llevarla a cabo en esta situación? ¿Cómo encontrar una solución nueva y útil en esta situación?



Este El estudio se ubica en la confluencia de un desafío operativo particular y reto operativo una tecnología y una perspectiva tecnológica. El caso estudio se centra en la empresa Puerto Lago Puno, una entidad representativa de la industria pesquera artesanal organizada del lago. Su funcionamiento, que se basa en el conocimiento tácito de su conocimiento pescadores de sus pescadores y en registros históricos que aún no se han analizado, es el ejemplo perfecto para probar la eficacia de una estrategia basada en datos y el efecto que produce. Los registros históricos que aún no se han analizado son el ejemplo perfecto para probar la eficacia de una estrategia basada en datos y el efecto que produce.

Esta tesis plantea la siguiente pregunta de investigación : ¿Cómo puede el diseño e implementación de un modelo predictivo que utiliza algoritmos de aprendizaje automático mejorar la eficiencia operativa de las operaciones de pesca artesanal dentro de la Compañía Puerto Lago Puno ? En respuesta, se establece un objetivo general claro: desarrollar y ejecutar un modelo predictivo que pronostique con precisión la cantidad de biomasa pesquera disponible , utilizando datos históricos de captura, meteorológicos y limnológicos . La hipótesis principal afirma que la aplicación de este modelo mejorará significativamente la eficiencia operativa, como lo indican indicadores como la Captura por Unidad de Esfuerzo (CPUE), la tasa de éxito de las tareas y la productividad de los insumos .

Este estudio se basa en un enfoque cuantitativo y en el diseño de modelado predictivo no experimental de una población con 7560 registros históricos. La metodología se organiza en un ciclo típico de ciencia de datos: 1) recopilación y preprocesamiento de datos; 2) ingeniería de características; 3) entrenamiento de algoritmos y validación comparativa (p. ej., Random Forest, XGBoost); y finalmente,



la evaluación de su impacto potencial. La relevancia del estudio es triple: a nivel práctico, para proporcionar a la empresa una herramienta de apoyo a la toma de decisiones que aumentará su rentabilidad; a nivel científico, para enriquecer la literatura relacionada con la aplicación de la IA en la pesca continental y sostenible; y a nivel social, para brindar mayor racionalidad en la gestión de uno de los recursos naturales más esenciales de Sudamérica.

Los siguientes capítulos se profundizarán en el marco teórico-conceptual, la metodología empleada, los resultados del modelado y el análisis de su impacto, para finalizar con conclusiones y recomendaciones específicas para la empresa y para investigaciones futuras en el área.

Los cinco capítulos en los que se dividen la tesis son los siguientes. En el Capítulo II se aborda el marco teórico con los fundamentos de la inteligencia artificial, el aprendizaje automático y las variables de predicción de capturas. Finalmente, las Conclusiones y Recomendaciones sintetizan los resultados, verifican las hipótesis y proponen lineamientos para su aplicación práctica y para futuras investigaciones.

CAPÍTULO I

ASPECTOS GENERALES

1.1. Descripción del problema

1.1.1 A nivel Internacional

A nivel mundial, la pesca artesanal se reconoce como un pilar de la seguridad alimentaria y el empleo en las comunidades costeras y ribereñas. Sin embargo, organizaciones como la FAO han documentado que este sector opera bajo una presión creciente, marcada por la sobreexplotación de los caladeros, los efectos del cambio climático y la competencia por los recursos. Una limitación transversal es la brecha tecnológica: mientras que la pesca industrial incorpora sistemas de monitoreo y análisis de datos, la pesca artesanal se basa principalmente en el conocimiento empírico y los métodos tradicionales. Esta dependencia genera un modelo de toma de decisiones reactivo, poco adaptado a la rápida variabilidad ambiental, que compromete tanto la productividad como la conservación a largo plazo de los ecosistemas. La falta de herramientas accesibles para predecir la biomasa disponible perpetúa un ciclo de incertidumbre e ineficiencia que afecta a millones de personas.

1.1.2 A nivel nacional

La pesca continental en el lago Titicaca representa una actividad socioeconómica relevante en el Altiplano peruano. Sin embargo, el análisis de grandes datos tiene sus limitaciones cuando las prácticas de gestión no respaldan herramientas predictivas eficaces. Instituciones como el IMARPE (Instituto Peruano de Investigación Marina) han descrito fluctuaciones preocupantes en la abundancia de especies de interés, como el carachi y el pejerrey, atribuibles a alteraciones en las condiciones limnológicas y climáticas. Si bien existe un marco regulatorio que promueve la pesca responsable, la transferencia de tecnología específica a operaciones de pesca artesanal es prácticamente inexistente. Esta deficiencia conduce a una planificación pesquera indiscriminada, donde el esfuerzo de explotación muestra poca correlación con la disponibilidad de recursos. Como resultado, se desperdician viajes de pesca y se gastan recursos innecesariamente, además de generarse una presión descoordinada sobre el ecosistema que impide la rentabilidad del sector y socava los objetivos más amplios orientados a la sostenibilidad.

1.1.3 A nivel local

Esta incertidumbre es una realidad cotidiana para la Compañía Puerto Lago Puno. Sus operaciones dependen de la pericia de sus pescadores, un conocimiento valioso que, por sí solo, no puede predecir el complejo comportamiento del lago. Cuentan con registros históricos de capturas y acceso a datos ambientales, pero ambos permanecen aislados, sin integrarse en un sistema que permita extraer patrones útiles para la planificación. El problema se centra en las operaciones pesqueras, que, basadas en estimaciones intuitivas, pueden resultar en bajos volúmenes de captura, lo que incrementa los costos de combustible y mano de obra

sin un retorno correspondiente. Además, sin pronósticos precisos, la logística posterior (transporte, procesamiento y comercialización) se gestiona de forma reactiva, lo que implica la pérdida de oportunidades de mercado. En otras palabras, se ven limitados por la falta de un modelo predictivo adaptado a su contexto, que les permita optimizar el uso de sus recursos, potenciar su productividad y establecer una ventaja competitiva basada en el conocimiento científico y técnico.

1.2. Formulación del problema

1.2.1 Problema Principal

¿De qué manera el diseño e implementación de un modelo predictivo basado en algoritmos de machine learning puede optimizar la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno en el 2024?

1.2.2 Problemas específicos

¿Qué datos históricos de captura de especies, volúmenes, zonas del Lago Titicaca están disponibles y son relevantes para las operaciones de la empresa?

¿Qué algoritmos de machine learning como Regresión Lineal, Random Forest, XGBoost o Redes Neuronales son los más adecuados para modelar la relación entre las variables ambientales y los volúmenes de captura?

¿Qué métricas de validación (Error Cuadrático Medio - MSE, Coeficiente de Determinación - R^2 , Error Absoluto Medio - MAE) demostrarán que el modelo tiene un rendimiento suficiente para su aplicación práctica?

¿En qué medida la implementación de un modelo predictivo basado en algoritmos de machine learning impacta en la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno?

1.3. Justificación de la investigación

1.3.1 Justificación Teórica

Esta investigación surge de una laguna de conocimiento previamente documentada en la literatura especializada. Evidencia teórica: a nivel teórico, existe abundante evidencia que relaciona las variables ambientales con la productividad pesquera en ecosistemas lacustres*. Sin embargo, hasta la fecha, no hemos encontrado trabajos previos sobre la implementación real de modelos predictivos de aprendizaje automático en la pesca artesanal de las tierras altas andinas y el lago Titicaca.

La mayoría de los modelos existentes se han aplicado a la pesca marina industrial o la acuicultura, donde las condiciones son más controlables. La alta dependencia de factores ecológicos dinámicos y del conocimiento tácito convierte a la pesca artesanal continental en un objeto de estudio específico. Este proyecto ampliará la base teórica adaptando algoritmos predictivos a un contexto de alta variabilidad con datos posiblemente limitados. Esto nos permitirá responder a algunas preguntas teóricas clave: ¿Qué clase de algoritmos (árboles de decisión, regresiones, redes neuronales) son más robustos en este entorno? ¿Cuál es la ponderación real de las variables limnológicas en comparación con las meteorológicas en este ecosistema? Datos hasta octubre de 2023. Los resultados contribuirán a la teoría sobre la idoneidad de la ciencia de datos para sistemas socioecológicos complejos y tradicionales.

1.3.2 Justificación práctico

La justificación práctica de esto es inmediata y concreta. La empresa Puerto Lago Puno opera en un entorno de incertidumbre que compromete su viabilidad

económica. Actualmente, las operaciones pesqueras se planifican con base en una combinación de experiencia colectiva e intuición, un enfoque útil pero, en última instancia, reactivo y propenso a errores.

Esta no es una investigación puramente académica; tiene objetivos utilitarios claros: crear una herramienta de apoyo a la toma de decisiones que transforme datos dispares en predicciones útiles. El modelo no pretende reemplazar el conocimiento tradicional del pescador, sino simplemente proporcionar información analítica adicional. En la práctica, este plan implicaría:

Reducción de costos operativos: Menos viajes a zonas con baja probabilidad de captura.

Mayor productividad: Concentrar los esfuerzos pesqueros en épocas y zonas con condiciones ambientales más favorables.

Mejor planificación logística: Proporcionar estimaciones de volúmenes para optimizar el transporte, el procesamiento y la venta.

Fomentar una pesca más sostenible: Prevenir la sobreexplotación de los recursos cuando son más vulnerables desde el punto de vista ecológico. Los cinco beneficios finales son de índole más económica y operativa: mayor eficiencia, mayor rentabilidad y una ventaja competitiva basada en la tecnología.

1.3.3 Justificación Metodológica

En el centro de este proyecto se encuentra una metodología de ciencia de datos que, a nuestro juicio, no solo facilita predicciones eficientes y aplicables, sino que también proporciona el rigor necesario mediante el empleo de un enfoque probado para la resolución de problemas de predicción con datos históricos. Justificación metodológica: Este marco permite una forma sistemática y replicable de convertir un problema empresarial mal definido en una solución técnica concreta.

El proceso propuesto —recopilación, preprocesamiento, que se gestionan de forma independiente y garantizan un modelado de calidad, junto con la validación y la evaluación— es lo que dicta la disciplina. La adopción de algoritmos de aprendizaje automático más sofisticados (en contraposición, por ejemplo, a modelos estadísticos lineales simples) tiene sentido, ya que capturan mejor las relaciones no lineales y complejas entre múltiples variables, que es precisamente lo que describe la interacción en la relación clima-lago-peces.

Además, la metodología incorpora una etapa de validación cruzada y evaluación mediante métricas objetivas (Error Cuadrático Medio - ECM o Coeficiente de Determinación - R^2) para garantizar que el modelo seleccionado no solo se ajuste bien a los datos de entrenamiento, sino que también se generalice correctamente a nuevos escenarios. Esta rigurosidad metodológica distingue al proyecto de las iniciativas de prueba de concepto, aportando credibilidad científica y demostrando su aplicabilidad práctica. La transparencia paso a paso también permitirá auditar el proceso y replicarlo o adaptarlo en el futuro.

1.4. Objetivos

1.4.1 *Objetivo general*

Diseñar e implementar un modelo predictivo basado en algoritmos de machine learning que permita optimizar la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno

1.4.2 *Objetivos específicos*

1. Recopilar y preprocesar datos históricos de captura, condiciones meteorológicas y datos limnológicos del Lago Titicaca relevantes para las operaciones de la empresa Puerto Lago Puno

2. Identificar y seleccionar los algoritmos predictivos más adecuados para modelar la relación entre las variables ambientales y los volúmenes de captura artesanales
3. Desarrollar, entrenar y validar el modelo predictivo con mayor rendimiento en la estimación de biomasa pesquera para apoyar la planificación de las faenas de pesca artesanal de la empresa
4. Evaluar el impacto del modelo predictivo en la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno

1.5. Importancia y alcance

Este estudio va más allá de los aspectos metodológicos de un modelo predictivo, ya que su relevancia radica en tres pilares clave:

Primero, el estudio ofrece una herramienta tangible para operacionalizar la toma de decisiones en la empresa a nivel económico y operativo.

Puerto Lago Puno. Se espera que lo anterior reduzca directamente los costos relacionados con viajes no rentables y, además, mejore la utilización de los insumos (combustible, mano de obra) al permitir la planificación del trabajo basada en pronósticos objetivos. Esto se traduce en mayores ganancias y una mejor competitividad, un ejemplo de cómo la tecnología accesible puede tener un impacto real en una industria tradicional.

En segundo lugar, la relevancia metodológica científica radica en que integra diversas variables —limnológicas, meteorológicas y datos históricos de capturas— para un ecosistema específico como el lago Titicaca. El modelo también generará nuevos conocimientos sobre las relaciones cuantitativas entre estos factores y la biomasa pesquera local, contribuyendo a subsanar una brecha de información en

la pesca continental andina. El enfoque adoptado podría servir de guía para iniciativas similares en otros lagos.

En tercer lugar, el proyecto fusiona los intereses comerciales con la sostenibilidad desde una perspectiva de gestión social y sostenible. Esperamos, por lo tanto, posibilitar una explotación más racional del recurso y evitar la sobrepesca durante los periodos de vulnerabilidad de las especies mediante un modelo predictivo preciso. Su contribución a la eficiencia de todos los procesos repercute directamente en la seguridad laboral de los pescadores y en la sostenibilidad a largo plazo de una actividad que forma parte del patrimonio cultural y económico de la región de Puno.

1.6. Limitaciones y delimitaciones

Esto se abordará de manera declarativa con transparencia metodológica, reconociendo que el estudio presenta limitaciones y restricciones inevitables:

Disponibilidad y calidad de los datos históricos: El rendimiento del modelo puede depender críticamente de la exhaustividad, la coherencia y el contexto de los historiales de captura de datos de la empresa. Cabe destacar que si las series de datos que abarcan los registros están incompletas, es decir, si existen interrupciones o errores en su registro, esto limitará la capacidad de entrenamiento y generalización de los algoritmos.

Actividad de estela completa: Acceso a datos ambientales en tiempo real. El modelo predictivo solo es útil operativamente si existe un flujo constante de datos meteorológicos (p. ej., vientos locales, humedad relativa, etc.) y limnológicos (propiedades físicas como la temperatura del agua). De hecho, la frecuencia de las predicciones dependerá de la disponibilidad y el costo de acceso a estas fuentes

de información (p. ej., estaciones SENAMHI o sensores locales).

El modelo predictivo es una probabilidad: Es importante destacar que los modelos de aprendizaje automático se basan en estimaciones y probabilidades basadas en patrones históricos. En algunos casos, factores inesperados o no cuantificables (como perturbaciones ecológicas repentinas o cambios en el comportamiento de las especies) pueden influir en las predicciones del modelo.

Factores humanos y de aceptación: La corrección técnica no es suficiente para garantizar el éxito del proyecto; su adopción es necesaria por parte de los pescadores y el gerente de la empresa. El uso efectivo inicial puede verse limitado por la resistencia al cambio o la desconfianza hacia una nueva herramienta, lo cual también se tendrá en cuenta, pero su gestión no se incluye en el alcance de este desarrollo tecnológico. Se definen los siguientes límites explícitos para garantizar la viabilidad y el enfoque del estudio.

Alcance espacial: El estudio se centra exclusivamente en las actividades pesqueras de la empresa Puerto Lago Puno en el lago Titicaca. No abarcará otras zonas del lago ni a empresas pesqueras rivales.

Delimitación temporal: El análisis de datos históricos se extenderá desde enero de 2019 hasta diciembre de 2023, según la serie completa proporcionada por la empresa. A lo largo de 2024 se llevará a cabo la fase de desarrollo y validación del modelo.

Delimitación de variables: Se utilizará un número limitado de variables predictoras, priorizando aquellas con disponibilidad comprobada y respaldo bibliográfico que justifique su influencia en la actividad pesquera del lago Titicaca. Estas incluirán:

Variables de captura: Volumen (kg) por especie principal (carachi, pejerrey), por día y zona.

Variables meteorológicas: Temperatura ambiente, velocidad/dirección del viento, precipitación.

Variables limnológicas: Temperatura de la superficie del agua, pH, oxígeno disuelto (sujeto a disponibilidad).

Delimitación tecnológica: El proyecto abarcará el diseño, entrenamiento y validación de un modelo predictivo de aprendizaje automático supervisado (los posibles modelos incluyen Bosque Aleatorio, Gradient Boosting o redes neuronales simples). Esto no incluye el desarrollo efectivo de hardware (sensores personalizados) ni el despliegue a gran escala de una plataforma de producción con conectividad en tiempo real; sin embargo, propondremos y evaluaremos un prototipo funcional.

1.7. Hipótesis

1.7.1 *Hipótesis General*

La implementación de un modelo predictivo basado en algoritmos de machine learning optimiza significativamente la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno en el 2024.

1.7.2 *Hipótesis Específicas*

1. El preprocesamiento de datos históricos de captura mediante técnicas de imputación y normalización permite construir un conjunto de datos robusto y confiable para el entrenamiento de los algoritmos predictivos
2. El algoritmo de Random Forest se identifica como el más adecuado, al presentar los mayores valores de precisión y menor error en la modelación de

la relación entre las variables ambientales y los volúmenes de captura

3. El modelo predictivo desarrollado y validado con el algoritmo seleccionado alcanza un nivel de precisión satisfactorio (ej. $R^2 > 0.75$) en la estimación de la biomasa pesquera
4. La implementación de un modelo predictivo basado en algoritmos de machine learning incrementa significativamente la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno

1.8. Variables e indicadores

1.8.1 Conceptualización de las variables *Variable 1: Algoritmo*

Modelo Predictivo

Es una herramienta computacional, producto de la aplicación de algoritmos predictivos, que se implementa para generar predicciones sobre la biomasa pesquera y las zonas óptimas de captura.

Variable 2: Eficiencia en la pesca artesanal.

Es el resultado operativo deseado que se busca optimizar. Se refiere a la productividad y efectividad del proceso de captura, medida a través de la maximización del rendimiento y la minimización de costos y esfuerzo.

CAPÍTULO II

FUNDAMENTOS TEÓRICOS

2.1. Antecedentes

2.1.1 Internacionales.

Un estudio realizado por Basurto y colaboradores (2023) bajo la iniciativa de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO), la Universidad de Duke y WorldFish, cuantificó el aporte global de la pesca artesanal o de pequeña escala. La investigación, que involucró a más de 800 personas en 58 países, determinó que esta actividad representa al menos el 40% de las capturas mundiales, contribuye a la seguridad alimentaria de 2.300 millones de personas y emplea a unos 60 millones de trabajadores, de los cuales el 40% son mujeres que participan tanto en la captura como en el procesamiento y comercialización. El estudio empleó una metodología de revisión integral de datos globales y estudios de caso para "iluminar las cosechas desconocidas", concluyendo que, a pesar de su colosal importancia socioeconómica y nutricional, el sector artesanal permanece frecuentemente invisible para los diseñadores de políticas públicas internacionales, que priorizan a la pesca industrial. Los investigadores argumentaron que esta falta de reconocimiento y apoyo institucional deja a las comunidades pesqueras artesanales en una situación de vulnerabilidad,

marginación y desventaja competitiva, limitando su capacidad para operar de manera sostenible y equitativa . Este antecedente es fundamental para la presente tesis porque establece el contexto global de relevancia y precariedad del sector artesanal, justificando la necesidad de investigaciones aplicadas, como el desarrollo de herramientas predictivas accesibles, que puedan empoderar a estos actores y mejorar su eficiencia operativa desde una base tecnológica sólida y adaptada a su realidad.

La investigación liderada por Xavier Basurto (2023) profundizó en las dinámicas de gobernanza y exclusión que enfrenta la pesca artesanal a nivel mundial. El estudio destacó que, si bien los pescadores artesanales poseen un conocimiento profundo de sus ecosistemas, frecuentemente operan sin derechos formales de participación en la gestión de los recursos pesqueros, incluso en territorios donde son habitantes originarios. Esta exclusión de los procesos de toma de decisiones se agrava por la marginalización geográfica y social de sus comunidades, donde suele haber una grave falta de servicios básicos como salud y educación, haciendo de la pesca la única opción económica viable. El trabajo utilizó análisis comparativo de políticas y entrevistas cualitativas para concluir que el dominio de la pesca industrial en el diseño normativo internacional y nacional deja a los pescadores artesanales en un estado de desamparo institucional, sin incentivos ni apoyo para adoptar prácticas más sostenibles, a pesar de su potencial natural para la conservación . Este antecedente es crucial para el presente proyecto, ya que subraya la importancia de desarrollar soluciones tecnológicas, como los algoritmos predictivos, que no solo sean herramientas operativas, sino también mecanismos para generar evidencia y empoderamiento local. Al dotar a la Empresa Puerto Lago Puno de una capacidad de análisis de datos propia, el

modelo puede servir como un instrumento para una gestión más informada y una participación más efectiva en los espacios de decisión sobre el recurso pesquero del Titicaca.

Basurto et al. (2023) demostraron que las mujeres desempeñan un papel marginal en la cadena de valor de la pesca artesanal. Si bien muchas mujeres pescan, la mayoría se dedica a actividades posteriores a la pesca, como la limpieza, el fileteado y la comercialización, a menudo omitidas por definiciones restrictivas de "pesca". Esta exclusión provoca que sus contribuciones económicas se subestimen y limita su acceso a beneficios sociales, capacitación o crédito. Mediante una perspectiva de género interseccional, el estudio ilustró cómo se realizan estas tareas a través de encuestas y relatos etnográficos.

Esta introducción es importante para la tesis, ya que resalta la necesidad de un enfoque más integral para la evaluación del impacto. El objetivo general de este proyecto es mejorar la eficiencia operativa a lo largo de toda la cadena poscosecha, y no solo durante la cosecha, como se ha evaluado. Un modelo predictivo optimizará los volúmenes y planificará el procesamiento y la comercialización. Este proceso, esencial para evaluar los impactos en cada fase de la implementación del modelo, está directamente vinculado con una mayor productividad y mejores condiciones laborales para las mujeres empleadas.

La evidencia internacional subraya la importancia de las tecnologías de pesca sostenible y las inversiones cooperativas para el futuro de la pesca artesanal (Basurto, 2023). La sostenibilidad requiere un enfoque holístico que abarque las dimensiones ecológica, económica y social. La investigación sienta las bases para determinar las herramientas tecnológicas necesarias para resolver los problemas existentes en el sector e indica que la sostenibilidad exige intervenciones

multifacéticas con el empoderamiento de las comunidades locales.

Este marco conceptual, indispensable para la investigación actual, se aplica directamente a las investigaciones donde el uso de algoritmos predictivos responde a la recomendación de invertir en tecnología sostenible. Sugiere que la inteligencia artificial puede mejorar la eficiencia económica y la optimización ecológica en la empresa Puerto Lago Puno, tal como lo demuestra este ejemplo concreto de recomendación internacional.

Revisión internacional con comparaciones regionales para la política pesquera: México, un caso singular; éxito documentado en modelos de cogestión basados en el conocimiento local y la gestión comunitaria. Sin embargo, las barreras burocráticas y la falta de representación siguen siendo obstáculos. Esta disertación contextualiza una propuesta técnica para Puerto Lago Puno como herramienta de gestión para buscar el diálogo con los organismos reguladores de la pesca peruana, incluido el IMARPE, y los efectos colaterales relacionados con la representación en el estilo de gobernanza adoptado por los decisores, quienes a veces privan a la pesca artesanal de presencia.

2.1.2 Nacionales

El Lago Titicaca, compartido por Perú y Bolivia, constituye el cuerpo de agua navegable más alto del mundo y un ecosistema de altísima biodiversidad para el altiplano andino. Investigaciones ictiológicas, como las sintetizadas en revisiones sobre la pesca en la región, estiman la presencia de 533 especies acuáticas, de las cuales 64 son endémicas y 23 son exclusivas de este lago. La ictiofauna nativa se centra en especies del género *Orestias* (conocidas como carachis), así como bagres de los géneros *Astroblepus* y *Trichomycterus*, las cuales han sido la base

proteica y cultural de las comunidades ribereñas desde tiempos precolombinos. Este antecedente es fundamental para la tesis porque define el objeto de estudio biológico y ecosistémico. El modelo predictivo que se desarrollará no operará en un vacío, sino en un entorno lacustre específico y complejo. La identificación de las especies nativas como principal objetivo de la pesca artesanal local (carachi, boga, suche, mauri) delimita las variables objetivo del modelo, mientras que la comprensión de su estatus de endemidad y biodiversidad subraya la importancia de una gestión pesquera precisa y sostenible, que el proyecto busca facilitar a través de la predicción.

Un antecedente crítico a nivel regional es la grave declinación de las especies nativas del Titicaca debido a múltiples presiones antrópicas. Estudios recientes, citados en análisis del estado del lago, advierten que hasta el 85% de las especies nativas han disminuido drásticamente. Entre los factores identificados se encuentran la sobrepesca, la contaminación por el crecimiento urbano e industrial, los impactos del cambio climático sobre la temperatura y el nivel del agua, y, de manera destacada, la competencia con especies exóticas introducidas. Esta problemática establece la urgencia y pertinencia de la investigación. El desarrollo de un modelo predictivo para la Empresa Puerto Lago Puno se enmarca directamente en la búsqueda de soluciones a la "sobrepesca" y a la "presión pesquera indiscriminada" mencionadas. Al permitir una planificación más eficiente que optimice el esfuerzo de captura (menos viajes infructuosos, mejores zonas y momentos), el modelo podría contribuir a reducir la presión innecesaria sobre el recurso, alineando la operación de la empresa con los urgentes objetivos de conservación de la biodiversidad nativa del lago.

La introducción histórica de especies exóticas, principalmente durante el

siglo XX, es un factor que ha reconfigurado por completo la dinámica pesquera del Titicaca y representa un antecedente ecológico ineludible. Con el objetivo de fomentar la pesca comercial, se introdujeron la trucha arcoíris (*Oncorhynchus mykiss*) y el pejerrey (*Odontesthes bonariensis*). Estas especies, más grandes y de rápido crecimiento, compitieron exitosamente con las nativas por alimento y espacio, alterando el equilibrio ecológico. Se atribuye, por ejemplo, que la presión de las especies introducida fue un factor clave en la desaparición del *Orestias cuvieri*. Para el modelo predictivo, este antecedente implica que la "biomasa pesquera" a estimar ya no es un conjunto homogéneo, sino una comunidad mixta de especies nativas e introducidas con interacciones complejas. El modelo deberá ser capaz, en la medida de lo posible, de diferenciar o ponderar las variables ambientales que afectan de manera distinta a estos grupos. La presencia del pejerrey como una de las principales especies comerciales hoy en día obliga a considerarlo dentro de las variables objetivo del modelo, a pesar de su origen exótico.

En el ámbito de las técnicas y conocimientos locales, la pesca artesanal en el Titicaca posee una rica tradición y un corpus de conocimiento ecológico tradicional (CET) transmitido por generaciones. Los pescadores de comunidades Uru y Aymara emplean embarcaciones tradicionales de totora y madera, y han desarrollado sofisticados sistemas de predicción basados en indicadores naturales. Por ejemplo, la presencia de gaviotas (qillwa en aymara) o de pequeños crustáceos y hongos llamados kankani en la superficie del agua son interpretados como señales de aglomeraciones de peces, guiando la selección del lugar de pesca. Este conocimiento, documentado etnográficamente, se complementa con una profunda familiaridad con las características batimétricas y dinámicas del lago en sus zonas

de operación . Este precedente es de suma importancia para el diseño metodológico de la tesis. El modelo predictivo basado en aprendizaje automático no debe considerarse un sustituto de este conocimiento ancestral, sino un complemento sistemático cuantitativo. Las variables que los pescadores ya observan (la presencia de aves, aspectos estacionales) pueden y deben incorporarse como características o variables predictivas en el conjunto de datos. La validación del modelo probablemente requerirá validación cruzada y se basará en el saber hacer local, una síntesis entre la inteligencia ecológica tradicional y la inteligencia artificial que busque sistemas de apoyo a la toma de decisiones más sólidos y culturalmente apropiados.

A nivel de políticas y manejo, existen diversas iniciativas de conservación y manejo sostenible para el Titicaca impulsadas por organizaciones nacionales e internacionales. Entre ellas se mencionan programas de repoblación con alevines de especies nativas, aunque los resultados específicos no se detallan ampliamente en la fuente consultada . Este antecedente, aunque general, sitúa el proyecto dentro de un contexto de esfuerzos por la sostenibilidad. La tesis puede posicionar al modelo predictivo como una herramienta de "manejo adaptativo" de precisión, que podría, a futuro, proporcionar datos valiosos para afinar este tipo de programas de repoblación. Al generar predicciones de biomasa por zona y tiempo, la empresa podría colaborar de manera más efectiva con las autoridades pesqueras, proporcionando datos verificados que informen dónde y cuándo realizar las siembras de alevines para maximizar su supervivencia y contribución al stock pesquero, cerrando así un ciclo virtuoso entre la actividad extractiva y la recuperación del recurso.

2.1.3 Locales

La Empresa Puerto Lago Puno se desenvuelve en un contexto operativo definido por la pesca artesanal de tipo comunitario y organizado. Como se describe en las prácticas del lago, los pescadores suelen organizar sus faenas con antelación, realizando reuniones para acordar la localización, especies objetivo e instrumentos a emplear. Los grupos, conformados típicamente por hasta quince integrantes con sus respectivas embarcaciones, operan con un conocimiento colectivo del territorio lacustre. Este antecedente operativo es la base sobre la cual se implementará el modelo. Define al usuario final: no un pescador individual, sino una empresa o colectivo que toma decisiones logísticas y estratégicas de manera grupal. El modelo predictivo debe diseñarse para apoyar precisamente este tipo de decisiones pre-faena (¿dónde ir? ¿cuándo ir? ¿qué esperar capturar?), optimizando la planificación del grupo en su conjunto. La factibilidad de implementación se ve favorecida por esta estructura organizativa, que puede facilitar la recopilación sistemática de datos de captura y la adopción coordinada de la nueva herramienta.

Un antecedente logístico clave es el uso de técnicas de pesca de arrastre adaptadas al contexto del Titicaca. Esta modalidad, descrita como una actividad que consta de dos embarcaciones que remolcan una red en forma de bolsa (con paños de redes agalleras), requiere una coordinación precisa y un esfuerzo considerable. Los pescadores arrean a los peces hacia la red desde una distancia, lo que demanda un buen conocimiento de la conducta de los peces y de la topografía del fondo. Esta técnica tiene implicaciones directas para el modelo. En primer lugar, representa un costo operativo significativo (combustible, desgaste de redes, esfuerzo humano) que justifica plenamente la búsqueda de eficiencia a través de la predicción. Un día de faena infructuosa implica un gasto de recursos considerable. En segundo lugar, la técnica misma sugiere variables predictoras

relevantes: la batimetría (profundidad) y el tipo de fondo en las zonas de pesca pueden ser factores críticos para el éxito del arrastre, por lo que deberían considerarse para su inclusión en el conjunto de datos, de estar disponibles.

La dependencia de indicadores naturales y conocimiento experiencial constituye el principal sistema de predicción actual de la empresa. Los pescadores, como se cita, basan sus decisiones en la observación directa (gaviotas, kankani) y en la experiencia acumulada: "constantemente vamos yendo a pescar, pescando nos damos cuenta si está en profundo o no tanto y como a qué lado ir... para tener buena pesca" . Este método, aunque valioso, es intrínsecamente cualitativo, subjetivo y difícil de escalar o transmitir. Además, es reactivo: los pescadores insisten en una zona si no hay captura inicial, pero esto puede llevar a perder tiempo en zonas realmente pobres. Este antecedente define el problema específico que la tesis busca resolver. El modelo predictivo propuesto ofrecería un sistema cuantitativo, objetivo y proactivo, capaz de analizar simultáneamente un historial largo de datos (capturas pasadas, condiciones meteorológicas históricas) para generar una recomendación probabilística antes de zarpar, complementando y potenciando el conocimiento experiencial con análisis de datos.

El ciclo de toma de decisiones y tolerancia al riesgo en la pesca local está bien descrito. Los pescadores manifiestan un proceso de prueba y error: si no consiguen buena pesca, insisten, y si persiste la mala racha, concluyen que "no llegó el pescado aún comer ahí" . Este comportamiento revela una alta tolerancia a la incertidumbre y a los resultados negativos consecutivos, pero también una ineficiencia inherente. El tiempo y recursos gastados en "insistir" en una zona pobre son pérdidas operativas. Este antecedente es crucial para la evaluación del impacto del modelo. Una métrica clave de éxito será precisamente la reducción en la

frecuencia y duración de estas "jornadas de insistencia" infructuosas. El modelo debería permitir a la empresa tomar la decisión de reubicarse con mayor rapidez y certeza, o mejor aún, evitar desde el principio las zonas con baja probabilidad pronosticada. El cambio cultural de un paradigma de "prueba y persistencia" a uno de "predicción y reasignación ágil" será un indicador tangible de la mejora en la eficiencia operativa.

Finalmente, el entorno socioeconómico y cultural de la pesca en el Titicaca, donde esta actividad es "vital como fuente de alimento" y "una actividad económica central", profundamente arraigada en festividades y rituales, proporciona el marco último de significación para la tesis. La Empresa Puerto Lago Puno no es solo una unidad productiva, sino un actor embebido en una red cultural y de subsistencia. Este antecedente recuerda que cualquier intervención tecnológica debe ser sensible al contexto. El modelo predictivo no es un fin en sí mismo, sino un medio para fortalecer esta actividad vital. Su éxito no se medirá solo en toneladas de pescado o soles ahorrados, sino también en su capacidad para ser adoptado de manera natural por los pescadores, respetando y dialogando con su conocimiento tradicional, y contribuyendo a la sostenibilidad a largo plazo de un modo de vida y una economía regional que depende del saludable equilibrio del Lago Titicaca.

2.2. Marco epistemológico

Este marco se fundamenta en una epistemología constructivista crítica. Parte de la idea de que el conocimiento sobre la eficacia de la pesca no es un reflejo directo de la realidad, sino más bien una construcción que surge de la interacción de datos empíricos (como las capturas y los factores ambientales), los actores (como los pescadores y las empresas) y los sistemas de interpretación (como los

algoritmos y el conocimiento tradicional).

Este enfoque es importante porque comprende que la pesca forma parte de sistemas sociales, económicos y ecológicos complejos, donde aspectos como la sostenibilidad, la equidad y el acceso al conocimiento tecnológico son fundamentales. El objetivo no es simplemente «descubrir» una relación causal oculta, sino desarrollar de forma colaborativa un instrumento predictivo válido y aplicable en este contexto particular, combinando el conocimiento local con la evaluación científica cuantitativa. La validez del modelo se evaluará no solo mediante métricas estadísticas (precisión, error), sino también por su eficacia práctica para mejorar la toma de decisiones y su significado cultural para los pescadores artesanales.

2.3. Estado del arte

La inteligencia artificial (IA) es un área de innovación tecnológica nueva y en rápido crecimiento que puede aplicarse a la industria pesquera internacional.

Los principales objetivos de las soluciones son la sostenibilidad, la eficiencia operativa y la gestión de recursos. Según la literatura especializada, la pesca industrial a gran escala se ha vuelto más común, y tecnologías como el aprendizaje automático y la visión artificial están transformando la forma en que se realizan las actividades. Un área de especialización es la predicción de biomasa y la identificación de los mejores caladeros. Empresas tecnológicas como la española Zunibal y diversos proyectos de investigación han desarrollado modelos predictivos que utilizan gran cantidad de datos históricos de captura por unidad de esfuerzo, junto con variables oceanográficas (temperatura de la superficie del mar, clorofila) y meteorológicas, para estimar la ubicación y cantidad de las poblaciones de peces.

Estos sistemas buscan encontrar las mejores rutas para los barcos, reducir el tiempo de búsqueda y disminuir el consumo de combustible y las emisiones de CO₂. Esto hace que cada viaje sea más rentable y respetuoso con el medio ambiente.

Simultáneamente, la IA se ha utilizado ampliamente en la optimización técnica de las operaciones y en la prevención de impactos ambientales no deseados. El diseño de los aparejos de pesca (por ejemplo, redes de arrastre y sus componentes) se modela mediante algoritmos de optimización o simulación para la ingeniería pesquera. Su objetivo es minimizar la resistencia del agua al capturar las especies objetivo, lo que se traduce en un ahorro significativo de combustible. Un avance de gran impacto en materia de sostenibilidad es el uso de la visión artificial para minimizar la captura incidental. Las especies pueden identificarse y clasificarse mediante modelos de detección de objetos en tiempo real (por ejemplo, YOLO (You Only Look Once)), que pueden integrarse en cámaras instaladas en las redes. Esto permite liberar de inmediato los ejemplares no deseados o de tamaño inferior al reglamentario, fomentando una temporada de pesca más ecológica y acorde con los criterios de sostenibilidad ambiental.

La IA también se sitúa a la vanguardia para garantizar el monitoreo, el control y la utilización sostenible de los recursos pesqueros, más allá de su uso en operaciones directas. Permite el monitoreo a escala global de la actividad de la flota pesquera mediante el análisis automatizado de imágenes satelitales, datos del Sistema de Identificación Automática (AIS) y otras fuentes de teledetección. Esta capacidad es fundamental para combatir la pesca ilegal, no declarada y no reglamentada (INDNR) en las Áreas Marinas Protegidas (AMP) y para hacer cumplir las regulaciones. Además, se crean sistemas de seguimiento basados en blockchain y análisis de datos para registrar permanentemente la fecha del

recorrido del pescado desde su captura hasta el consumidor, asegurando que el producto pesquero sea legal, de calidad y sostenible a lo largo de las cadenas de suministro.

Modelado microbiano en la era de la digitalización y la automatización: un panorama innovador. Este trabajo surge a raíz de revisiones recientes (2023) que identifican el estado actual de la técnica, pero con escasa investigación disponible sobre estudios que vayan más allá de los enfoques descriptivos. Los avances tecnológicos analizados se centran casi exclusivamente en el medio marino y la pesca industrial, donde existe infraestructura y capital para integrar sensores costosos, sistemas satelitales y capacidad de computación avanzada. El campo de las aplicaciones de IA dirigidas a la pesca artesanal continental carece claramente de atención, especialmente en ecosistemas acuáticos complejos como el lago Titicaca. Esta brecha es importante debido al contexto de la pesca artesanal —con sus embarcaciones pequeñas, acceso limitado a datos en tiempo real y alta dependencia de factores ambientales locales—, lo que requerirá soluciones tecnológicas de bajo costo y alta adaptabilidad que respeten e integren el conocimiento ecológico tradicional.

Por ello, es en este vacío donde encuentra el fundamento y la posibilidad de su aporte la presente tesis "Aplicación de algoritmos predictivos para mejorar las eficiencias operativas en la pesca artesanal de la empresa Puerto Lago Puno 2024". En lugar de recrear modelos contruidos para el océano, este proyecto busca innovar adaptando y validando enfoques de aprendizaje automático en el contexto único de la pesca artesanal lacustre. Su objetivo es crear un modelo matemático que prediga, a partir de datos locales y fácilmente obtenibles (historiales de capturas artesanales, condiciones meteorológicas en la región de Puno y

principales variables limnológicas del lago), pronósticos útiles para una planificación óptima de la pesca. De esta manera, la investigación propone una solución tecnológica contextualizada, práctica y escalable, que no solo proporciona beneficios en la eficiencia operativa y económica de la empresa, sino que también busca abordar uno de sus problemas con una gestión más informativa para una gestión más sostenible de los recursos pesqueros del lago Titicaca; por lo tanto, combinando un proceso innovador de análisis de datos con necesidades reales del sistema de producción tradicional que es vital para la región.

2.4. Bases teóricas

2.4.1 Fundamentos Conceptuales de la Inteligencia Artificial y el Aprendizaje Automático

El nacimiento formal de la Inteligencia Artificial (IA) tuvo lugar en la Conferencia de Dartmouth de 1956, organizada por John McCarthy, Marvin Minsky, Nathaniel Rochester y Claude Shannon.

Este fue el inicio de la IA como campo científico formal, con el objetivo de crear máquinas capaces de simular algunos aspectos de la inteligencia humana, como el aprendizaje y la resolución de problemas. Sin embargo, sus orígenes filosóficos y conceptuales se remontan mucho más atrás en el tiempo, desde los autómatas de la antigüedad y los razonamientos de pensadores como Descartes y Leibniz, hasta la formalización lógica por parte de figuras como George Boole y Gottlob Frege en los siglos XIX y XX, quienes desarrollaron estilos factorizados y fundamentos matemáticos para el razonamiento simbólico. La máquina de Turing (1936) estableció los límites de lo que se podía computar, mientras que la prueba de Turing de 1950 propuso un método para evaluar si una máquina era verdaderamente inteligente.

La IA ha progresado a través de ciclos de optimismo y desilusión (los "veranos de la IA" seguidos de limitaciones), pero su evolución reciente ha sido sorprendente. Hubo un invierno de la IA debido a los desafíos que enfrentaron los sistemas basados en reglas en las décadas de 1950 y 1960. Las redes neuronales experimentaron un resurgimiento en las décadas de 1980 y 1990, pero se vieron limitadas por la capacidad de procesamiento y los datos. El siglo XXI fue, sin duda, un punto de inflexión, ya que la confluencia del big data, el aumento de la capacidad de procesamiento y mejores algoritmos permitieron que la IA dominara tareas como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la evaluación de datos ambientales y operativos complejos.

2.4.2 Definición y alcance de la IA en el contexto contemporáneo

La IA, del Instituto de Ingenieros Eléctricos y Electrónicos (IEEE), es un subcampo interdisciplinario de la informática que busca construir sistemas que realicen funciones que normalmente requieren capacidad humana, como la percepción, la comprensión del lenguaje, el aprendizaje a partir de la experiencia y el razonamiento para la planificación y la organización. La inteligencia artificial (IA) es el estudio de agentes inteligentes que perciben estímulos ambientales a través de sus sensores y luego trabajan para alcanzar una meta, según Russell y Norvig (2021). La IA se divide en IA débil (o estrecha) e IA fuerte (o general). Actualmente, la mayor parte de la IA que nos rodea es IA débil (ANI), especializada en realizar tareas específicas como el reconocimiento facial, los sistemas de recomendación de productos, la traducción de idiomas y la predicción de series temporales.

La inteligencia artificial general (IAG) fuerte, capaz de emular el procesamiento cognitivo humano, sigue siendo una hipótesis. Actualmente, la IA se

utiliza en robótica, sistemas de recomendación, vehículos autónomos, diagnóstico médico y, en general, en análisis avanzados y optimización de procesos en sectores más convencionales como la pesca. La IA amplía las capacidades humanas al analizar interacciones complejas entre variables ambientales y operativas, lo que transforma la toma de decisiones, pasando de un enfoque empírico a uno basado en datos.

2.4.3 El Aprendizaje Automático (Machine Learning) como motor de la IA aplicada

El subcampo más vibrante de la IA es el Aprendizaje Automático (Machine Learning o ML). A diferencia de la programación tradicional, los sistemas ML aprenden patrones a partir de datos. Según Mitchell (1997) "Se dice que un programa informático aprende de la experiencia E con respecto a alguna clase de tareas T y una medida de rendimiento P si su rendimiento en una tarea en T , medido por P , mejora con la experiencia E ". El ML tiene como objetivo encontrar funciones matemáticas que relacionen variables de entrada con una variable de salida, minimizando el error de predicción.

El modelo de aprendizaje automático (ML) es el principal mecanismo que impulsa la inteligencia artificial aplicada, ya que aprende de los datos en lugar de depender de conocimientos codificados explícitamente, como la IA tradicional. Los algoritmos de ML pueden aprender patrones complejos en grandes conjuntos de datos. En la pesca artesanal, permite que los modelos se beneficien del historial de capturas, las condiciones climáticas y ambientales para descubrir reglas sobre la abundancia de peces. El ML transforma datos históricos desorganizados en un activo estratégico, proporcionando información predictiva y una ventaja competitiva; en esencia, pone la IA en acción para sectores que dependen de los recursos.

2.4.4 Tipos de aprendizaje: supervisado, no supervisado y por refuerzo

La teoría del Aprendizaje Automático se organiza en paradigmas según la naturaleza de la "experiencia" (E) de la que aprende el algoritmo. Los tres tipos fundamentales son:

Aprendizaje Supervisado (Supervised Learning): Es el paradigma más relevante para tareas de predicción y clasificación. En este esquema, el algoritmo se entrena con un conjunto de datos etiquetados, donde cada ejemplo de entrada viene acompañado de la salida correcta conocida (la "etiqueta" o label). El objetivo del modelo es aprender la función que mejor mapee las entradas a esas salidas conocidas, de manera que pueda predecir con precisión la etiqueta para nuevos datos nunca vistos. Para la predicción de capturas, esto implica tener un registro histórico donde, para cada día o faena (entrada con variables como temperatura, viento, etc.), se conoce el volumen real de pescado capturado (la etiqueta). El modelo aprenderá de estos pares entrada-salida. Se subdivide en:

Regresión: Cuando la variable a predecir es continua o numérica (ej.: kilogramos de captura, temperatura futura).

Clasificación: Cuando la variable a predecir es categórica o discreta (.: "alta captura", "media captura", "baja captura").

Aprendizaje No Supervisado (Unsupervised Learning): Aquí, los datos de entrenamiento no tienen etiquetas. El objetivo del algoritmo es descubrir la estructura interna, los patrones ocultos o las agrupaciones naturales dentro de los datos. Técnicas como el clustering (agrupamiento) o el análisis de componentes principales (PCA) son típicas. En el contexto pesquero, podría usarse para segmentar diferentes tipos de días de pesca según las condiciones ambientales, o

para identificar correlaciones no evidentes entre múltiples variables limnológicas, sin predecir un volumen específico.

Aprendizaje por Refuerzo (Reinforcement Learning - RL): En este paradigma, un agente aprende a interactuar con un entorno dinámico. El agente realiza acciones y recibe del entorno recompensas (positivas o negativas) como feedback. Su objetivo es aprender una política óptima (secuencia de acciones) que maximice la recompensa acumulada a largo plazo. Aunque más complejo, el RL tiene potencial para problemas de planificación y optimización secuencial a largo plazo en pesquerías, como la definición de estrategias de pesca sostenibles que maximicen el rendimiento sin agotar el stock.

Para el objetivo central de esta tesis la predicción de biomasa el aprendizaje supervisado, y específicamente los modelos de regresión, es el paradigma teórico y metodológico fundamental.

2.4.5 La Pesca de Precisión: Paradigma de aplicación de la IA en el sector

La Pesca de Precisión (PP) emerge como el paradigma conceptual que sintetiza la aplicación de tecnologías de la información, sensores, datos y análisis con la IA y el ML en su núcleo—para la gestión optimizada y sostenible de la actividad pesquera. Analogía directa de la "Agricultura de Precisión", la PP busca aplicar el principio de intervención correcta, en el lugar correcto, en el momento correcto y de la manera correcta. Su objetivo final es incrementar la eficiencia operativa y económica mientras se minimiza el impacto ambiental y se promueve la sostenibilidad del recurso.

Teóricamente, la PP se sustenta en el ciclo de observación-decisión- acción realimentado. La IA actúa como el sistema nervioso central de este ciclo:

Observación: Recopilación masiva y automatizada de datos del entorno (satélites, boyas, sensores a bordo), de la operación (GPS, sonares) y del recurso (capturas por especie y talla).

Análisis y Decisión: Procesamiento de estos datos mediante modelos de ML para generar conocimiento accionable: predicciones de zonas de abundancia, recomendaciones de rutas, alertas de condiciones adversas o estimaciones de captura incidental.

Acción: Implementación de las decisiones en el campo, cerrando el ciclo y generando nuevos datos para alimentar y mejorar los modelos.

La adopción de la PP representa una transición de un modelo de pesca reactivo y basado en experiencia, sujeto a una alta incertidumbre, hacia un modelo proactivo y basado en datos. En el contexto de la pesca artesanal del Lago Titicaca, este paradigma no implica necesariamente la adopción de tecnología satelital costosa, sino la adaptación de sus principios fundamentales al uso de datos disponibles (meteorológicos, limnológicos, históricos) y al desarrollo de modelos predictivos accesibles. Así, la empresa puede optimizar la planificación de sus faenas, reduciendo el esfuerzo de búsqueda y los costos operativos (combustible, tiempo), mientras contribuye a una explotación más racional del recurso, alineándose con los objetivos de la pesca sostenible. La PP es, por tanto, el marco de aplicación que da sentido y dirección a la implementación de algoritmos de ML en esta investigación.

2.4.6 Fundamentos del análisis predictivo

El análisis predictivo es la disciplina que utiliza técnicas estadísticas, de minería de datos y de aprendizaje automático para analizar datos históricos y actuales con el fin de hacer predicciones probabilísticas sobre eventos o

comportamientos futuros. Su fundamento teórico reside en la premisa de que los patrones identificables en el pasado y el presente pueden informar sobre lo que es probable que ocurra en el futuro, asumiendo cierta estabilidad en las relaciones subyacentes. No busca la certeza absoluta, sino cuantificar la incertidumbre y proporcionar la mejor estimación posible dados los datos disponibles.

En el núcleo del análisis predictivo está la construcción de un modelo predictivo, una representación matemática o computacional simplificada de la realidad. Este modelo captura la relación funcional entre un conjunto de variables predictoras (independientes o features) y una variable objetivo (dependiente o target) que se desea predecir. La relación se expresa generalmente como: $Y \approx f(X_1, X_2, \dots, X_n)$, donde Y es la variable objetivo (ej.: volumen de captura) y f es la función modelada que depende de las variables predictoras $X_1 \dots X_n$ (ej.: temperatura del agua, velocidad del viento). La calidad de la predicción depende críticamente de: 1) la calidad y relevancia de los datos de entrada, 2) la idoneidad del algoritmo elegido para capturar la naturaleza de la relación f (lineal, no lineal, compleja), y 3) la robustez de la validación del modelo con datos no vistos durante el entrenamiento.

2.4.7 Proceso de desarrollo de un modelo predictivo (From Data to Deployment)

El desarrollo de un modelo predictivo es un proceso iterativo y estructurado, a menudo representado como un ciclo de vida. Para esta investigación, el proceso teórico clave es el flujo de Aprendizaje Automático Supervisado, que consta de las siguientes fases:

Definición del Problema y Obtención de Datos: Se especifica claramente qué se va a predecir (la variable Y) y qué datos se necesitan. Se recopilan los datos históricos de las fuentes identificadas (empresa, SENAMHI, etc.).

Preprocesamiento y Limpieza de Datos (Data Wrangling): Fase crucial que consume la mayor parte del tiempo. Incluye la imputación de valores faltantes, la detección y tratamiento de valores atípicos (outliers), la normalización o estandarización de variables numéricas (para que estén en escalas comparables) y la codificación de variables categóricas (ej.: convertir "dirección del viento" en variables numéricas).

Ingeniería de Características (Feature Engineering): Como se detallará en la sección 3.4, es el proceso de crear, transformar o seleccionar las variables predictoras (X) más informativas para el modelo a partir de los datos crudos. Es donde el conocimiento del dominio (pesca, limnología) se incorpora al modelo.

División de los Datos: Los datos preparados se dividen aleatoriamente, típicamente en:

Conjunto de Entrenamiento (Training Set, ~70-80%): Para ajustar los parámetros del modelo.

Conjunto de Validación (Validation Set, ~10-15%): Para ajustar los hiperparámetros del modelo y comparar diferentes algoritmos durante el desarrollo.

Conjunto de Prueba (Test Set, ~10-15%): Para evaluar el rendimiento final del modelo elegido con datos completamente independientes y no vistos. Esta evaluación simula su desempeño en el mundo real.

Selección, Entrenamiento y Evaluación del Modelo: Se eligen uno o varios algoritmos candidatos (ej.: Regresión Lineal, Random Forest). Cada algoritmo se entrena (ajusta) utilizando el conjunto de entrenamiento para aprender la función f . Luego, se evalúa su rendimiento inicial en el conjunto de validación usando métricas apropiadas (ver 2.5). Este paso suele ser iterativo.

Ajuste de Hiperparámetros (Hyperparameter Tuning): Los algoritmos tienen configuraciones llamadas hiperparámetros (ej.: profundidad de un árbol, tasa de aprendizaje) que no se aprenden de los datos, sino que se fijan antes del entrenamiento. Se optimizan sistemáticamente (con técnicas como Grid Search o Random Search) usando el conjunto de validación para maximizar el rendimiento.

Validación y Despliegue Final: El modelo con la mejor configuración se evalúa una última vez en el conjunto de prueba, que ha permanecido totalmente aislado, para obtener una estimación imparcial de su error de generalización. Si el rendimiento es satisfactorio, el modelo se despliega en un entorno de producción (ej.: una aplicación web o script) para generar predicciones con nuevos datos. Finalmente, se establece un ciclo de monitoreo para reevaluar su desempeño a lo largo del tiempo y retrenarlo si es necesario (deriva del modelo).

2.4.8 Clasificación de modelos según su salida: regresión, clasificación y series de tiempo

Los modelos predictivos de ML se clasifican teóricamente según la naturaleza de la variable objetivo Y que predicen:

Modelos de Regresión: Predicen una variable numérica continua. Es la categoría directamente aplicable a la predicción de volúmenes de captura (kg). La salida del modelo es un número real. Ejemplos: predecir el precio de una casa, la temperatura o, en este caso, la biomasa capturable.

Modelos de Clasificación: Predicen una etiqueta o categoría discreta. La salida es una clase entre un conjunto finito de posibilidades. Puede ser binaria (ej.: "captura alta" vs. "captura baja") o multiclase (ej.: "alta", "media", "baja"). Si bien no es el enfoque principal de esta tesis, una estrategia híbrida podría ser útil: primero clasificar si el día será de alta/baja captura y luego, para los días de alta, usar un

modelo de regresión para estimar el volumen preciso.

Modelos de Series de Tiempo (Time Series): Un caso especial de regresión donde las observaciones se recogen en intervalos de tiempo secuenciales y equidistantes. La característica clave es la dependencia temporal (autocorrelación): el valor en el tiempo t depende de los valores en tiempos $t-1$, $t-2$,... Para predecir capturas, donde el orden temporal es crucial

(estacionalidad, tendencias), las técnicas de series de tiempo son altamente relevantes. Los modelos pueden ir desde ARIMA (clásicos) hasta redes neuronales recurrentes (LSTM) avanzadas. En la práctica, un problema de predicción de capturas puede abordarse tanto con modelos de regresión estándar (incluyendo la fecha como feature procesada) como con modelos específicos de series de tiempo.

2.4.9 Algoritmos de regresión para predicción numérica: Lineal, Ridge, Lasso y basados en árboles

La teoría del aprendizaje automático ofrece una amplia gama de algoritmos para regresión. La elección depende de la complejidad de los datos y la relación subyacente.

Regresión Lineal: Es el modelo más simple y fundamental. Asume una relación lineal entre las variables predictoras y la objetivo. Ajusta una línea (o hiperplano) que minimiza la suma de los errores cuadrados entre las predicciones y los valores reales. Su principal ventaja es la interpretabilidad: los coeficientes indican el cambio esperado en Y por un cambio unitario en X . Sin embargo, su supuesto de linealidad rara vez se cumple en fenómenos ambientales complejos, por lo que su utilidad directa aquí puede ser limitada, aunque sirve como baseline.

Regresión Ridge y Lasso (Regresión Lineal Penalizada): Son extensiones

de la regresión lineal que incorporan una penalización a la magnitud de los coeficientes del modelo. Esto reduce la varianza (sobreajuste) y mejora la generalización, especialmente cuando hay muchas variables predictoras correlacionadas.

Ridge (Regresión de Cresta): Penaliza la suma de los cuadrados de los coeficientes (norma L2). Tiende a reducir todos los coeficientes pero rara vez los lleva a cero exactamente.

Lasso (Least Absolute Shrinkage and Selection Operator): Penaliza la suma de los valores absolutos de los coeficientes (norma L1). Tiene la propiedad de llevar algunos coeficientes a cero exactamente, realizando así selección de variables automática. Esto es muy útil para identificar las variables ambientales más influyentes en la captura.

Algoritmos Basados en Árboles (Tree-Based): Son familias de modelos no lineales que no asumen una relación paramétrica específica entre variables. Son especialmente potentes para capturar interacciones complejas.

Árboles de Decisión (Decision Trees): Modelan decisiones y sus consecuencias mediante una estructura de árbol. Son muy interpretables, pero propensos al sobreajuste.

Random Forest (Bosques Aleatorios) para Regresión: Es un método de ensamblado (ensemble) que combina las predicciones de muchos árboles de decisión (un "bosque"), cada uno entrenado con una muestra aleatoria de los datos y un subconjunto aleatorio de las variables. La predicción final es el promedio de las predicciones de todos los árboles. Es robusto, maneja bien datos no lineales y reduce drásticamente el sobreajuste. Es uno de los algoritmos más utilizados y confiables para problemas de regresión en dominios aplicados como este.

Gradient Boosting Machines (GBM) - XGBoost, LightGBM: Otro potente método de ensamblado, pero secuencial. Entrena árboles uno tras otro, donde cada nuevo árbol corrige los errores residuales del anterior. Es conocido por su alta precisión predictiva, aunque puede ser más propenso al sobreajuste si no se regula adecuadamente. XGBoost (Extreme Gradient Boosting) es una implementación optimizada muy popular.

Para la predicción de capturas, donde las relaciones entre variables ambientales y biomasa son intrínsecamente no lineales y multivariadas, los algoritmos basados en árboles (Random Forest y Gradient Boosting) se presentan teóricamente como los candidatos más sólidos y prometedores, seguidos de la regresión Lasso como técnica para selección inicial de variables.

2.4.10 Métricas de evaluación del rendimiento predictivo

La evaluación teórica y cuantitativa de un modelo de regresión es esencial para determinar su utilidad práctica. No basta con que el modelo se ajuste bien a los datos de entrenamiento; debe generalizar bien a datos nuevos. Las métricas clave son:

Error Cuadrático Medio (Mean Squared Error - MSE): Es la métrica fundamental. Calcula el promedio de los cuadrados de las diferencias entre los valores predichos y los reales: $MSE = (1/n) * \sum (y_{real} - y_{pred})^2$. Al elevar al cuadrado los errores, penaliza más los errores grandes. Su unidad es el cuadrado de la unidad original (ej.: kg²), lo que a veces dificulta su interpretación directa.

Raíz del Error Cuadrático Medio (Root Mean Squared Error - RMSE): Es simplemente la raíz cuadrada del MSE: $RMSE = \sqrt{MSE}$. Tiene la misma unidad que la variable original (kg), por lo que es más intuitiva. Un RMSE de 50 kg significa

que, en promedio, las predicciones del modelo se desvían unos 50 kg del valor real. Es la métrica de error absoluto más utilizada.

Error Absoluto Medio (Mean Absolute Error - MAE): Calcula el promedio de los valores absolutos de los errores: $MAE = (1/n) * \sum |y_{real} - y_{pred}|$. A diferencia del RMSE, no penaliza desproporcionadamente los errores grandes, dando una visión más lineal del error promedio.

Coeficiente de Determinación (R^2 o R-cuadrado): Mide la proporción de la varianza en la variable objetivo que es explicada por el modelo. Varía entre 0 (el modelo no explica nada) y 1 (el modelo explica toda la variabilidad). Un R^2 de 0.75 significa que el modelo captura el 75% de la variación observada en las capturas. Es una excelente métrica de ajuste relativo, pero debe interpretarse junto con el RMSE/MAE para tener una visión completa del error absoluto.

En la práctica, durante el entrenamiento y validación, se debe minimizar el RMSE (o MAE) y maximizar el R^2 en el conjunto de prueba. Un modelo con un R^2 alto (ej., >0.8) y un RMSE bajo en relación con el rango típico de capturas, indicaría una capacidad predictiva sólida y útil para la planificación operativa de la empresa.

2.4.11 Variables y dimensiones para la predicción de capturas

Esta dimensión constituye el núcleo de lo que el modelo pretende predecir o explicar la variable objetivo (Y). Su definición operativa y calidad son críticas.

Volumen de Captura (Biomasa): La variable principal es el peso total (en kilogramos) de pescado capturado por faena, por día o por zona. Este es el indicador más directo de éxito productivo. Se debe desagregar, idealmente, por especie objetivo principal (carachi, pejerrey), ya que las diferentes especies pueden responder de manera distinta a las condiciones ambientales. Los datos históricos

de la empresa son la fuente primaria.

Indicadores Derivados de Eficiencia: Además del volumen bruto, se pueden construir variables objetivo más sofisticadas que capturen mejor el concepto de "eficiencia operativa", alineándose con el objetivo general de la tesis:

Captura por Unidad de Esfuerzo (CPUE): Es un indicador estándar en ciencia pesquera. Se calcula como el volumen de captura (kg) dividido por una medida del esfuerzo aplicado. En pesca artesanal, el esfuerzo puede cuantificarse como: número de embarcaciones por faena, horas totales de pesca, o consumo de combustible por faena. La CPUE (ej., kg/embarcación, kg/hora) es una métrica mucho más robusta que el volumen bruto, ya que normaliza la captura por el recurso invertido, permitiendo comparaciones más justas entre días o estrategias.

Rendimiento Económico Estimado: Combinando el volumen de captura por especie con precios de mercado históricos, se puede generar una variable objetivo de valor económico de la captura (en soles). Esto alinea directamente la predicción del modelo con la rentabilidad de la empresa.

La elección de la variable objetivo (volumen bruto vs. CPUE) tendrá implicaciones en el modelo. Predecir CPUE puede ser teóricamente más desafiante pero también más valioso, ya que optimizar la eficiencia (captura por recurso gastado) suele ser el objetivo último de la empresa.

2.4.12 Variables predictoras ambientales - factores externos determinantes

Esta dimensión abarca las variables del entorno físico que influyen teóricamente en la fisiología, distribución y comportamiento de los peces, y por ende, en su disponibilidad para la captura. Son las variables predictoras

(X) clave.

Variables Meteorológicas (Atmosféricas):

Temperatura del Aire: Afecta indirectamente la temperatura superficial del agua y la actividad metabólica de los peces. Fuente: SENAMHI.

Precipitación (lluvia): La escorrentía puede alterar la turbiedad, el nivel del lago y la entrada de nutrientes, afectando las cadenas tróficas.

Velocidad y Dirección del Viento: El viento genera oleaje y mezcla de las capas de agua (upwelling), lo que puede concentrar nutrientes y plancton, atrayendo a los peces. La dirección puede indicar el movimiento de masas de agua.

Presión Atmosférica y Humedad Relativa: Menos estudiadas, pero cambios bruscos en presión se asocian a cambios en el comportamiento de los peces.

Variables Limnológicas (del Agua del Lago): Son, teóricamente, las más directamente vinculadas a la productividad acuática. Su medición suele ser más compleja.

Temperatura del Agua: Es el factor abiótico más importante. Cada especie tiene un rango óptimo de temperatura para su alimentación y actividad. Cambios de incluso 1-2°C pueden afectar significativamente la distribución.

Nivel de Oxígeno Disuelto (OD): Los peces requieren niveles mínimos de OD para respirar. En aguas cálidas o estancadas, el OD puede disminuir, forzando a los peces a migrar a zonas más oxigenadas.

pH (Acidez/Alcalinidad): El Lago Titicaca es naturalmente alcalino.

Variaciones fuera del rango normal pueden causar estrés fisiológico.

Transparencia/Turbidez (Secchi): Indica la cantidad de material en suspensión (plancton, sedimentos), relacionada con la productividad primaria

(alimento para los peces) y la visibilidad.

Conductividad/Salinidad y Nutrientes (Nitratos, Fosfatos): Indicadores de la fertilidad del lago y base de la cadena alimenticia.

Variables de Contexto Operativo y Espacial - El Factor Humano y Geográfico

Esta dimensión incorpora factores no ambientales que teóricamente influyen en el resultado de la captura.

Variables Temporales (Features de Tiempo): La temporalidad es crucial en ecología.

Componentes Cíclicos: Se extraen de la fecha. Incluyen el día del año (para capturar la estacionalidad anual), el mes, la semana del año, y componentes cíclicos como $\sin(2\pi * \text{día}/365)$ y $\cos(2\pi * \text{día}/365)$ para modelar estacionalidad de forma suave.

Tendencia: Un componente que captura cambios a largo plazo (ej., disminución gradual de stocks).

Día de la semana/Festivo: Puede afectar la intensidad de la actividad pesquera.

2.4.13 Variables espaciales y batimétricas:

Zona de Pesca (Georreferenciación): La ubicación (coordenadas GPS aproximadas de la faena) es fundamental. El lago tiene zonas de distinta productividad. Se puede codificar como categoría (ej., "Bahía de Puno", "Cerca de las Islas de los Uros") o usar las coordenadas directamente (latitud, longitud) si hay suficientes datos.

Profundidad (Batimetría): Diferentes especies prefieren diferentes rangos de

profundidad, que a su vez se correlacionan con temperatura, luz y disponibilidad de alimento. Es una variable predictora potencialmente muy poderosa si se tiene acceso a un mapa batimétrico o estimaciones de profundidad por zona.

2.4.14 Variables Operativas y de Esfuerzo:

Tipo de Arte de Pesca: Si la empresa usa más de un método (red de cerco, red de arrastre ligera, línea).

Tamaño de la Tripulación / Número de Embarcaciones: Una proxy directa del esfuerzo de pesca aplicado, esencial si la variable objetivo es el volumen bruto y no la CPUE.

2.4.15 Ingeniería de Características: Integración y Transformación de Variables Crudas

La Ingeniería de Características (Feature Engineering) es el proceso teórico y creativo de transformar las variables predictoras crudas (dimensiones 2 y 3) en características (features) que sean más informativas y adecuadas para que los algoritmos de ML descubran patrones. Es donde el conocimiento del dominio se convierte en ventaja algorítmica. Operaciones clave incluyen:

Creación de Features Derivadas:

Promedios Móviles (Moving Averages): Calcular la temperatura promedio de los últimos 3, 5 o 7 días. Los peces pueden responder a condiciones promedio, no solo a las del día exacto.

Retrasos Temporales (Lags): Incluir el valor de una variable (ej., temperatura) de 1, 2 o 3 días antes. Esto captura la inercia de los sistemas ecológicos.

Interacciones: Crear nuevas variables como el producto o cociente entre

dos.Temperatura_Agua * Oxígeno_Disuelto (índice de estrés), Precipitación / Viento (índice de mezcla).

Variables de Cambio: Diferencia entre el valor de hoy y el de ayer para una variable (ej., cambio en temperatura).

Codificación de Variables Categóricas: Transformar variables no numéricas (ej.: "Zona_A", "Zona_B", "Dirección_Norte") en números. Técnicas comunes son One-Hot Encoding (crear una columna binaria por categoría) o Target Encoding (sustituir la categoría por el promedio histórico de captura en esa categoría).

Escalado/Normalización: Llevar todas las variables numéricas a una escala común (ej., media=0, desviación=1) para que algoritmos sensibles a la escala (como Regresión Lineal, SVM) funcionen correctamente. Los algoritmos basados en árboles (Random Forest) no requieren este paso.

Selección de Características (Feature Selection): Identificar y retener solo las variables más relevantes para la predicción, eliminando ruido y reduciendo la complejidad del modelo. Técnicas como la importancia de características (proporcionada por Random Forest) o los coeficientes de Regresión Lasso son instrumentalmente teóricas para este fin.

2.5. Marco conceptual

Algoritmo de Aprendizaje Automático (Machine Learning Algorithm):

Conjunto de reglas y procedimientos matemáticos definidos paso a paso que permite a un sistema informático identificar patrones en los datos. A través de un proceso de entrenamiento, el algoritmo ajusta automáticamente sus parámetros internos para mejorar su desempeño en una tarea específica, como la clasificación o la predicción numérica (regresión).

Aprendizaje Supervisado (Supervised Learning):

Paradigma fundamental del aprendizaje automático en el cual un algoritmo se entrena utilizando un conjunto de datos etiquetados. Cada ejemplo en estos datos está compuesto por un vector de características (variables de entrada) y su correspondiente valor de salida conocido (la "etiqueta" o variable objetivo). El objetivo del algoritmo es aprender una función que mapee las entradas a las salidas, para luego hacer predicciones precisas sobre datos nuevos no vistos.

Captura por Unidad de Esfuerzo (CPUE):

Indicador estándar en ciencia pesquera que mide la eficiencia de la actividad extractiva. Se calcula como el volumen de captura (en peso o número de individuos) dividido por una medida cuantificable del esfuerzo pesquero aplicado para obtenerla. Las unidades comunes son kilogramos por hora de pesca (kg/h) o kilogramos por embarcación (kg/emb). La CPUE se utiliza como un proxy de la abundancia relativa del recurso y es una métrica central de eficiencia operativa.

Conjunto de Entrenamiento (Training Set):

Un conjunto de datos históricos para entrenar los parámetros del modelo de predicción. El conjunto que se utiliza para entrenar el algoritmo, que generalmente abarca entre el 70 % y el 80 % de los datos.

Conjunto de Prueba (Test Set):

Existen datos fuera de muestra que se reservan al comienzo de la investigación y permanecen aislados mientras se lleva a cabo el entrenamiento y ajuste del modelo, para proporcionar una medida de rendimiento imparcial para información nueva y no vista con respecto a la utilizada para la construcción/entrenamiento/ajuste.

Eficiencia Operativa Pesquera:

Grado de optimización en el uso de los recursos productivos (tiempo, combustible, mano de obra, embarcaciones) para maximizar el rendimiento (captura) de la actividad pesquera. Se evalúa mediante indicadores como la CPUE, la tasa de éxito de las faenas (porcentaje de salidas con captura significativa) y la productividad de los insumos (ej., kg de pescado por litro de combustible).

Error Cuadrático Medio (MSE) y su Raíz (RMSE):

Las métricas más importantes para evaluar la precisión de un modelo de regresión son el MSE y el RMSE. El error cuadrático medio (MSE) calcula el promedio de las diferencias al cuadrado entre los valores predichos y los valores reales. El RMSE, que es la raíz cuadrada compuesta, proporciona la métrica en unidades originales (como kilogramos) para una interpretación más sencilla, en lugar de elevarla al cuadrado. Cuanto menor sea el RMSE, mejor y más preciso será el modelo.

Ingeniería de Características (Feature Engineering):

La ciencia de datos (o ingeniería de características, como se la conoce comúnmente) se refiere esencialmente al proceso de transformar datos brutos en nuevas variables predictoras (características), que a su vez pueden agregarse a los algoritmos para obtener mayor poder predictivo. Estas pueden ser, por ejemplo, promedios móviles o interacciones con la variable.

Inteligencia Artificial (IA):

Inteligencia Artificial (IA): Subcampo de la informática que se ocupa del diseño de sistemas de hardware y software para realizar tareas que normalmente requieren inteligencia humana, como la percepción visual, el reconocimiento de voz

o la toma de decisiones y la traducción entre idiomas. Actualmente, busca resolver problemas específicos mediante el uso de datos.

Límnología:

Ciencia interdisciplinaria que estudia los ecosistemas acuáticos continentales (lagos, lagunas, ríos, humedales), sus propiedades físicas (temperatura, corrientes), químicas (oxígeno, pH, nutrientes) y biológicas, así como sus interacciones. Proporciona un marco teórico para evaluar cómo las condiciones del agua en el lago Titicaca influyen en las poblaciones de peces.

Modelo Predictivo (Predictive Model):

Las relaciones entre las variables predictoras y la variable objetivo se representan en un modelo de aprendizaje automático. Una vez validado, este modelo utiliza nuevos datos de entrada para predecir valores futuros de la variable objetivo.

Pesca Artesanal: La pesca artesanal, que se realiza en pequeñas embarcaciones con tecnología sencilla y a menudo emplea mano de obra familiar o local, está ligada a las comunidades costeras o ribereñas, donde posee un valor sociocultural. En el lago Titicaca, se refiere principalmente a pequeñas embarcaciones y redes.

Pesca de Precisión (Precision Fishing):

Un nuevo sistema de gestión pesquera basado en la agricultura de precisión que combina tecnologías de la información, GPS, sensores y análisis de datos (incluidos modelos de IA) con la toma de decisiones para optimizar las decisiones. En esencia, busca maximizar los esfuerzos pesqueros de manera eficiente, selectiva y sostenible manteniendo la incertidumbre sobre dónde y cómo pescar

con un mínimo en `method.log`.

Preprocesamiento de Datos (Data Preprocessing):

La limpieza y transformación de datos es una etapa importante del ciclo de la ciencia de datos. Incluye la limpieza de los datos brutos para su posterior análisis, lo que implica completar los valores faltantes, corregir errores, regularizar las escalas a 0 y 1, y codificar las variables categóricas.

Random Forest (Bosque Aleatorio):

Algoritmo de aprendizaje automático supervisado, perteneciente a la familia de métodos de ensamblado (ensemble). Opera construyendo una multitud de árboles de decisión durante el entrenamiento. La predicción final para un problema de regresión es el promedio de las predicciones de todos los árboles individuales. Es reconocido por su robustez, alta precisión, capacidad para manejar datos no lineales y por proporcionar un ranking de la importancia de cada variable predictora.

Regresión (en Machine Learning):

Tipo de tarea de aprendizaje supervisado en la cual la variable objetivo que el modelo debe predecir es una cantidad numérica continua. A diferencia de la clasificación (que predice una etiqueta discreta), la regresión estima un valor numérico. En esta investigación, el volumen de captura (kg) o la CPUE son variables de regresión.



CAPÍTULO III

METODOLOGÍA Y RESULTADOS

3.1. Enfoque de la investigación

Esta investigación se enmarca en un enfoque cuantitativo. Este enfoque se caracteriza por la medición numérica de variables previamente definidas, el uso de técnicas estadísticas y de modelado matemático para analizar relaciones entre ellas, y la búsqueda de generalizaciones a partir de patrones observados en una muestra de datos. Esta es la mejor manera de proceder, ya que el objetivo principal es crear y probar un modelo predictivo, lo cual requiere mediciones precisas, pruebas de hipótesis estadísticas y evaluación del rendimiento mediante métricas numéricas objetivas como el error cuadrático medio y el coeficiente de determinación R^2 . El conocimiento se construye a partir de evidencia empírica derivada de datos históricos, con el fin de establecer relaciones causales o predictivas basadas en correlaciones y patrones cuantificables.

3.2. Métodos aplicados de investigación

La investigación se fundamenta en la combinación de dos metodologías científicas sinérgicas: Método hipotético-deductivo: Esta es la principal forma de estructurar el trabajo. Parte de una hipótesis teórica general basada en el marco

teórico y el estado del arte: «Existe una relación funcional predecible entre un conjunto específico de variables ambientales y contextuales y el volumen o la eficiencia de las capturas de la pesca artesanal en el lago Titicaca, la cual puede modelarse mediante algoritmos de aprendizaje automático». A partir de esta hipótesis general, se formulan hipótesis operativas específicas y comprobables, por ejemplo: «El algoritmo Random Forest alcanzará un coeficiente de determinación (R^2) superior a 0,75 en la predicción del volumen de captura en el conjunto de prueba». Estas hipótesis se validan empíricamente mediante experimentación computacional (entrenamiento y evaluación del modelo). Los resultados respaldan o refutan las hipótesis, lo que permite extraer conclusiones sobre la capacidad predictiva del sistema.

Método Analítico-Sintético: Se utiliza de forma transversal en todo el ciclo de ciencia de datos. La fase de análisis implica descomponer el problema en sus partes: observar las diferentes dimensiones de los datos (captura, meteorología, limnología), observar la distribución de cada variable y observar las correlaciones bivariadas. La fase de síntesis implica juntar esas partes en un todo nuevo y más complejo: el modelo predictivo final, que combina las relaciones entre todas las variables para hacer una única predicción. Este método es fundamental en la fase de Ingeniería de características, donde se examinan variables individuales para posteriormente sintetizar nuevas que sean más informativas.

3.3. Tipo de investigación

La investigación es de carácter tecnológico aplicado. Es aplicada porque su objetivo principal es solucionar un problema específico y útil: la ineficiencia en la planificación de las operaciones de pesca artesanal. Busca generar tanto conocimiento teórico como una solución tecnológica práctica. Es tecnológica

porque el producto principal es un prototipo de software (el modelo predictivo) desarrollado mediante ciencia de datos e inteligencia artificial en un área específica. Su objetivo es mejorar la toma de decisiones de la empresa mediante una herramienta que utiliza datos para apoyarla.

3.4. Nivel

El estudio se sitúa en el nivel explicativo-predictivo. En su fase intermedia, busca determinar y medir cómo las diferentes variables predictoras (como el entorno o el espacio) influyen en la variable de resultado (captura). Para ello, utiliza algoritmos y técnicas como el análisis de importancia de características, con el fin de identificar los factores más relevantes. Su enfoque es fundamentalmente predictivo, ya que su objetivo principal es anticipar un evento o magnitud futura (el volumen o la eficiencia de la captura) a partir del modelo matemático establecido. Este es el nivel óptimo para alcanzar el objetivo general de diseñar un modelo que optimice la eficiencia, lo que requiere no solo comprender relaciones, sino anticipar resultados.

3.5. Diseño

La investigación adopta un diseño no experimental, transversal y de modelado predictivo.

No Experimental: Porque no se manipulan ni controlan variables de forma activa en un entorno de campo. El investigador observa, analiza y modela datos preexistentes (los 7,560 registros históricos) sin intervenir en el proceso real de pesca. Se estudian las variables tal como ocurrieron naturalmente.

Transversal (Transeccional): Porque los datos analizados corresponden a un periodo de tiempo ya ocurrido (ej., 2018-2023), capturando una "instantánea"

temporal que incluye múltiples eventos y ciclos. Los datos se recolectan y analizan en un solo momento para el estudio, aunque representen información histórica.

Diseño de Modelado Predictivo (Predictive Modelling Design): Es el diseño específico dentro de la ciencia de datos. Sigue un protocolo estandarizado: 1) preparación de datos, 2) división en entrenamiento/validación/prueba, 3) entrenamiento iterativo de algoritmos, 4) selección y ajuste del mejor modelo, y 5) evaluación final con datos de prueba. Este diseño está orientado exclusivamente a la construcción y validación de un instrumento de pronóstico, evaluando su precisión y capacidad de generalización.

3.6. Ámbito de investigación

Alcance de la investigación: Define los límites del estudio en términos de limitaciones espaciales, temporales, conceptuales y metodológicas. Es importante para mantener el rumbo, el enfoque y la claridad, evitar la dispersión y delimitar las conclusiones. Este manifiesto se articula en torno a cuatro dimensiones: espacial, temporal, disciplinaria y orientada a la acción.

Ámbito Espacial

La investigación se limita geográficamente al área de operaciones pesqueras de la empresa Puerto Lago Puno en el lago Titicaca. Esto incluye específicamente las áreas del lago donde la empresa realiza habitualmente sus actividades de pesca artesanal y para las cuales posee registros históricos de capturas. No abarca la totalidad del lago binacional, ni a otras empresas pesqueras o comunidades de la región. La delimitación espacial precisa se definirá mediante coordenadas geográficas o sectores nominales (por ejemplo, bahía interior, área cercana a islas específicas) derivados de los datos históricos de la empresa. Esta delimitación es



crucial para que el modelo predictivo aprenda de las condiciones ambientales específicas y los patrones de captura de las áreas donde opera la empresa, garantizando así la relevancia y la aplicabilidad directa de sus pronósticos.

Ámbito Temporal

El estudio abarca dos periodos temporales diferenciados:

Los datos históricos de captura y las variables ambientales se obtendrán de un período de tiempo significativo, que dependerá de los registros completos de la empresa; por ejemplo, de enero de 2019 a diciembre de 2023. Este lapso de cinco ciclos abarcará patrones estacionales y variabilidad para un mejor ajuste del modelo.

Fase de descubrimiento, desarrollo (creación), prueba y validación del modelo predictivo: Diseñaremos, desarrollaremos/entrenaremos, validaremos y evaluaremos el modelo predictivo en 2024, y las predicciones se evaluarán durante esta fase..

Ámbito Disciplinar y Conceptual

La investigación abarca tres campos:

Aprendizaje automático, específicamente aprendizaje automático supervisado (modelos de regresión y predicción); Ciencia de datos e Inteligencia artificial aplicada.

Disciplinas: Ciencias pesqueras y limnología, que orientan la comprensión teórica de las variables ambientales y la dinámica de las poblaciones de peces en los ecosistemas lacustres, en el contexto de la pesca de precisión.

Gestión de Operaciones y Eficiencia Empresarial: La solución tecnológica se contextualiza dentro de un objetivo empresarial: mejorar la eficiencia operativa,

medida mediante indicadores como la captura por unidad de esfuerzo (CPUE), la productividad de los insumos y el desempeño económico.

La investigación no profundiza en aspectos de biología pesquera avanzada (por ejemplo, modelos de dinámica poblacional con parámetros de crecimiento y mortalidad) ni en la ingeniería de hardware para la captura automatizada de datos in situ. Su enfoque se centra en el desarrollo de un modelo de software predictivo basado en datos fácilmente disponibles o de acceso relativamente sencillo.

Ámbito de Acción (Alcance de la Intervención)

La intervención que proponemos consiste en investigación y desarrollo aplicados al prototipo, centrándonos en:

Resultado: El estudio se basará en datos históricos existentes y no implementará nuevos sensores para recopilar datos en tiempo real, aunque se podría recomendar su viabilidad futura.

Proceso: Los datos se utilizarán para crear, entrenar y validar/predecir un modelo de aprendizaje automático que posteriormente se evaluará (resultado) y mostrará las capacidades predictivas/del prototipo (por ejemplo, una interfaz de usuario que genere predicciones para nuevos datos).

Salida (Output) y Resultado (Outcome): El producto final es el modelo validado y la evaluación de su potencial impacto sobre los indicadores de eficiencia definidos. La investigación no incluye la implementación a gran escala, el despliegue en producción continua ni el mantenimiento operativo a largo plazo del sistema dentro de la empresa, aunque proveerá las especificaciones técnicas para ello. Su éxito se medirá por la precisión predictiva del prototipo y la evidencia sólida

de su utilidad potencial para la toma de decisiones.

3.7. Población y muestra

3.7.1 Población

La población de la presente investigación está constituida por datos históricos y no por individuos, en concordancia con su naturaleza cuantitativa y el enfoque de ciencia de datos. La población objetivo la conforman todos los registros históricos de las operaciones de pesca artesanal de la Empresa Puerto Lago Puno, así como los datos ambientales asociados al Lago Titicaca que potencialmente influyen en la productividad pesquera.

El marco poblacional abarca 7,560 registros correspondientes al periodo, los cuales fueron obtenidos y consolidados a partir de fuentes primarias de la empresa y secundarias de instituciones oficiales. Esta población de datos se compone de tres dimensiones interrelacionadas:

Datos de Captura: Incluyen variables como fecha, zona de pesca (georreferenciada), especie objetivo, volumen de captura (kg) y esfuerzo pesquero (horas/embarcación).

Datos Meteorológicos: Comprenden variables como temperatura ambiental, velocidad del viento, precipitación y horas de sol, obtenidos del SENAMHI.

Datos Limnológicos: Incluyen parámetros de la calidad del agua como temperatura superficial, pH y transparencia, proporcionados por la Autoridad Binacional del Lago Titicaca (ALT).

3.7.2 Muestra

Dado que la investigación se basa en el análisis de datos existentes y el desarrollo de un modelo predictivo, no se empleó una técnica de muestreo

probabilístico tradicional. En su lugar, se trabajó con el universo completo de 7,560 registros disponibles después de los procesos de recolección, integración y depuración de datos.

Para el entrenamiento y evaluación de los algoritmos de machine learning, este conjunto de datos completo se dividió de forma aleatoria en tres subconjuntos:

Conjunto de Entrenamiento (Training Set): Comprendió el 70% (5,292 registros) del total. Su función fue permitir que los algoritmos aprendieran los patrones y relaciones subyacentes entre las variables.

Conjunto de Validación (Validation Set): Representó el 15% (1,134 registros). Se utilizó para fine-tunar los hiperparámetros de los modelos y para la selección del algoritmo con mejor desempeño durante la fase de desarrollo.

Conjunto de Prueba (Test Set): Constituyó el 15% restante (1,134 registros). Este conjunto se mantuvo completamente aislado y fue utilizado únicamente en la etapa final para evaluar el rendimiento y la capacidad de generalización del modelo definitivo con datos nunca vistos, proporcionando así una medida imparcial de su eficacia predictiva.

3.8. Métodos y técnicas de recogida de información

3.8.1 Técnicas de investigación

Este estudio emplea la experimentación computacional como metodología principal dentro del marco cuantitativo. Este método implica el diseño, la ejecución y el análisis de experimentos sistemáticos realizados exclusivamente en un entorno de software, utilizando datos históricos para construir, evaluar y verificar modelos matemáticos. En el aprendizaje automático, la experimentación computacional se lleva a cabo mediante el pipeline de ciencia de datos, un proceso estructurado e



iterativo que transforma los datos brutos en un modelo predictivo funcional.

El uso de este método en la investigación se basa en la idea de realizar pruebas y validaciones empíricas controladas. Una hipótesis principal plantea la existencia de una relación funcional predecible entre un conjunto de variables ambientales/operativas (X) y el volumen o la eficiencia de la captura (Y). Un algoritmo de aprendizaje automático f describe esta relación, de modo que $Y \approx f(X)$. El método implica probar rigurosamente varios algoritmos candidatos para la función f , medir su error de predicción y seleccionar el que genere el menor error en datos no vistos previamente. Esto demuestra la hipótesis de que las predicciones son posibles.

Las fases operativas de esta técnica, que estructuran el procedimiento metodológico, son:

Preprocesamiento y exploración de datos: Limpieza, modificación y uso de estadísticas descriptivas para conocer la calidad y distribución de los datos poblacionales.

Ingeniería de características: Creación y selección de las mejores variables predictoras (características) a partir de los datos brutos.

Modelado y entrenamiento: Uso repetido de algoritmos de aprendizaje supervisado en el conjunto de entrenamiento para aprender la función f .

Validación y evaluación: Utilizando los conjuntos de validación y prueba, medimos el rendimiento de los modelos candidatos con métricas cuantitativas como RMSE y R^2 .

Análisis de resultados: Comprender la capacidad predictiva del modelo final y la importancia relativa de las variables utilizadas.

Este método es repetible, objetivo y permite controlar las variables dividiendo

los datos en grupos, lo que garantiza que las conclusiones sobre el rendimiento del modelo se basen en datos medibles.

3.8.2 Instrumento de investigación

Una plataforma de programación y análisis computacional es la herramienta principal para recopilar y, sobre todo, analizar datos en este estudio. En lugar de utilizar cuestionarios o entrevistas convencionales, se emplea un entorno de software especializado que permite modificar 7560 registros de población, ejecutar algoritmos complejos y automatizar los procesos de modelado y evaluación.

Herramienta principal: Entorno de programación Python con bibliotecas especializadas

Utilizamos Python (versión 3.9 o superior) debido a su versatilidad y ecosistema de bibliotecas, especialmente para aplicaciones científicas y de ciencia de datos. Esta herramienta integral permite la ejecución de todas las fases de la técnica. Los componentes clave incluyen:

Bibliotecas de manipulación y análisis de datos:

Pandas → Carga, limpieza y transformación de DataFrames (manejo de hasta 7560 registros, fusión de tablas, etc.)

NumPy: Proporciona soporte para operaciones matemáticas eficientes en matrices multidimensionales.

Bibliotecas de visualización:

Gráficos estáticos: Matplotlib y Seaborn: Crean diagramas de dispersión, histogramas, etc., esenciales para la exploración y transcripción de resultados.

mgl: Bibliotecas de aprendizaje automático y modelos predictivos.

La herramienta principal para experimentar de forma computacionalmente eficiente es skytlearn (o sklearn). Cuenta con una excelente documentación e



implementaciones bien definidas:

Algoritmos de preprocesamiento: Escaladores (StandardScaler, MinMaxScaler), imputadores de datos faltantes (SimpleImputer).

Algoritmos de modelado: Regresión Lineal, Ridge, Lasso, Random Forest, Gradient Boosting (XGBoost, si se instala), entre otros.

Herramientas de validación: Funciones para dividir datos (train_test_split), validación cruzada (cross_val_score), búsqueda de hiperparámetros (GridSearchCV, RandomizedSearchCV).

Métricas de evaluación: Cálculo de RMSE, MAE, R^2 (mean_squared_error, r2_score).

Entorno de Desarrollo Integrado (IDE):

Jupyter Notebook o VS Code con una extensión de Python es el laboratorio experimental que nos permite ejecutar nuestro código de forma interactiva, documentar cada paso del análisis, visualizar los resultados y capturar un flujo de trabajo reproducible. El último cuaderno es un registro de ese experimento.

Herramientas secundarias:

Sistema de control de versiones (VCS) (Git): Permite realizar un seguimiento del código fuente del proyecto, incluyendo el historial y las copias de seguridad.

Herramientas de oficina (Microsoft Excel/Google Sheets, Word): Para procesar los datos brutos y redactar el informe final.

3.8.3 Criterios de Inclusión

Los criterios de inclusión definen las condiciones mínimas que debe cumplir un registro dentro de la población total de 7,560 para ser considerado válido y ser incorporado al conjunto de datos de desarrollo del modelo. Estos criterios garantizan la integridad, relevancia y coherencia mínima de los datos.

Integridad de la Variable Objetivo: Se incluirán todos los registros donde la variable dependiente principal (volumen de captura en kg o CPUE calculada) esté completamente registrada y sea un número positivo mayor a cero. Los registros con captura nula o fallida (valor 0) se analizarán por separado, pero se mantendrán si representan información válida sobre el esfuerzo sin éxito.

Disponibilidad de Variables Clave Predictoras: Se incluirán registros que cuenten, como mínimo, con datos para un núcleo de variables predictoras críticas consideradas teóricamente esenciales. Este núcleo incluirá: fecha de la faena, temperatura ambiental (como proxy básico de la energía del sistema) y una identificación de la zona de pesca (nominal o coordenada). La falta de datos en otras variables se resolverá mediante técnicas de imputación.

Consistencia Temporal: Se incluirán registros cuyas fechas estén dentro del rango temporal definido y validado (ej., 2018-2023) y que sean lógicamente consistentes (ej., la fecha de captura no puede ser posterior a la fecha de registro en el sistema).

Correspondencia Espaciotemporal: Se incluirán registros donde los datos meteorológicos y limnológicos puedan ser asociados de manera confiable a la fecha y zona de la faena. Por ejemplo, un registro de captura en la Bahía de Puno debe poder vincularse con los datos ambientales medidos para esa misma zona o la estación meteorológica más cercana en esa fecha.

3.8.4 Criterios de Exclusión

Los criterios de exclusión identifican registros que, por su naturaleza o calidad, introducirían ruido, sesgo o error en el modelo y, por tanto, deben ser eliminados del conjunto de desarrollo principal.

Registros de Prueba o Erróneos: Se excluirán registros claramente

identificados como datos de prueba, duplicados exactos, o entradas con valores imposibles (ej., volumen de captura de 15,000 kg en una faena artesanal, coordenadas GPS fuera de los límites del Lago Titicaca).

Valores Atípicos Irrecuperables (Outliers): Se excluirán registros con valores atípicos extremos y multivariados que, tras un análisis estadístico (ej., basado en la distancia de Mahalanobis o reglas del rango intercuartílico para múltiples variables), se determine que son producto de un error de medición o registro, y no de un fenómeno ecológico real. Los valores atípicos univariados que sean biológicamente posibles (ej., una captura excepcionalmente alta en un día) se conservarán.

Faenas con Condiciones Excepcionales No Modelables: Se excluirán registros correspondientes a eventos disruptivos extraordinarios que no representan las condiciones operativas normales y que el modelo no podrá predecir. Esto incluye faenas realizadas durante paros laborales, desastres naturales agudos (como una tormenta severa no registrada en los datos estándar) o periodos de veda oficial estricta donde la actividad reportada es incongruente.

Variables con Datos Faltantes Críticos y No Imputables: Se excluirán registros donde falten más del 50% de las variables predictoras consideradas en el modelo final, o donde falten todas las variables de una dimensión completa (ej., todos los datos limnológicos). La imputación en estos casos introduciría más incertidumbre que información.

3.9. Validez y confiabilidad del instrumento de investigación

3.9.1 Validación de los instrumentos

En una investigación basada en ciencia de datos, la "validación del instrumento" se refiere a garantizar que la plataforma de software y los algoritmos utilizados miden y modelan correctamente lo que se pretende.

Validez de Contenido (Validez Teórica): El instrumento (el pipeline de código en Python) es válido en contenido porque:

Las librerías (scikit-learn, pandas) implementan algoritmos matemáticos estándar y ampliamente publicados en la literatura de machine learning y estadística.

La selección de algoritmos (Random Forest, Gradient Boosting, etc.) se basa en el marco teórico que sostiene que son adecuados para capturar relaciones no lineales propias de sistemas ecológicos.

Se utilizan métricas de evaluación estandarizadas (RMSE, R^2) aceptadas por la comunidad científica para medir el error de predicción en regresión.

Validez de Criterio (Validez Práctica/Predictiva): La validez más importante del instrumento será empírica y se demostrará al final del proceso: Validez

Concurrente: El modelo se validará comparando sus predicciones con los datos reales del Conjunto de Prueba, que han sido aislados. Una alta correlación y un bajo error (RMSE) entre las predicciones y la realidad constituyen la prueba de su validez predictiva.

El proceso de validación cruzada (k-fold cross-validation) durante el entrenamiento sirve como validación interna y recurrente del instrumento, asegurando que el rendimiento no es producto del azar o de un ajuste a un subconjunto específico de datos.

3.9.2 Confiabilidad de los instrumentos

La confiabilidad se refiere a la estabilidad, precisión y reproducibilidad de los instrumentos de software.

Confiabilidad de la Medición (Precisión del Código): Las funciones en librerías como scikit-learn son deterministas y de alta precisión numérica. Un mismo



algoritmo, con los mismos datos y los mismos hiperparámetros (random_state fija), producirá exactamente los mismos resultados cada vez que se ejecute el código. Esto elimina la variabilidad por error del instrumento.

Confiabilidad por Consistencia Interna (Reproducibilidad): La confiabilidad se garantiza mediante:

Uso de Semillas Aleatorias (Random State): Todas las funciones que involucren aleatoriedad (división de datos, algoritmos basados en árboles) utilizarán una semilla fija (ej., random_state=42). Esto garantiza que cualquier investigador, en cualquier máquina, pueda reproducir exactamente el mismo modelo partiendo de los mismos datos crudos y el código fuente.

Documentación del Pipeline de Código: Todo el proceso, desde la carga de datos hasta la predicción final, quedará documentado en un Jupyter Notebook o script comentado. Este documento es el protocolo que asegura la consistencia y transparencia del procedimiento.

Confiabilidad por Consistencia Externa (Estándares de la Industria): La confiabilidad está respaldada por la adopción masiva de estas herramientas. Python y scikit-learn son estándares de facto en investigación académica e industrial, sujetos a revisiones de código continuas, lo que asegura que son instrumentos robustos, depurados y confiables para la tarea..



CAPÍTULO IV

RESULTADOS

4.1. Presentación, análisis e interpretación de datos

El presente capítulo constituye el núcleo empírico de la investigación, donde se procesan, analizan e interpretan los datos para responder a los objetivos planteados. La exposición sigue una estructura lógica que refleja el flujo de trabajo del modelado predictivo: primero, se describe la preparación del conjunto de datos; segundo, se presentan los resultados del entrenamiento y la selección del algoritmo óptimo; tercero, se valida el modelo final; y finalmente, se analizan los factores que más influyen en las predicciones. Todo el proceso se ejecutó utilizando el entorno de programación Python 3.9 y las librerías especializadas Pandas, Scikit-learn, XGBoost y Matplotlib/Seaborn para la visualización.

Figura 1*Estructura del DataFrame*

```
<class
'pandas.core.frame.DataFrame'> RangeIndex: 7560
entries, 0 to 7559 Data
columns (total 9 columns):
#      Column                                Non-Null Count  Dtype
-----
0      fecha                                7560 non-null     object
1      zona_pesca                          7560 non-null     object
2      especie                             7560 non-null     object
3      temperatura_agua_c                  7560 non-null     float64
4      oxigeno_disuelto_mg_l               7560 non-null     float64
5      ph_agua                             7560 non-null     float64
6      velocidad_viento_kmh                7560 non-null     float64
7      precipitacion_mm                    7560 non-null     float64
8      volumen_captura_kg                  7560 non-
null float64 dtypes: float64(6),
object(3)
memory usage: 531.7+ KB

<class 'pandas.core.frame.DataFrame'>
```

Indica que los datos están almacenados en un DataFrame de la librería Pandas, una estructura tabular bidimensional similar a una hoja de cálculo, utilizada para el análisis y procesamiento de datos.

El conjunto de datos analizado está conformado por 7,560 registros históricos de pesca artesanal y 9 variables, de las cuales seis son de tipo numérico continuo y tres de tipo categórico. Todas las variables presentan valores completos, sin datos faltantes, lo que garantiza la calidad del dataset para el entrenamiento y validación de modelos predictivos basados en algoritmos de Machine Learning

Figura 2*Selección de variables*

```
X =
df.drop(columns=["volumen_captura_kg",
"fecha"]) y = df["volumen_captura_kg"]

categoricas = ["zona_pesca",
"especie"] numericas = [
    "temperatura_agua_c"
,

```

```
"oxigeno_disuelto_mg  
_l", "ph_agua",  
"velocidad_viento_km  
h",  
"precipitacion_mm"
```

Este fragmento de código define las variables predictoras (X) y la variable objetivo (y), así como la clasificación de las variables según su naturaleza (categóricas y numéricas), paso fundamental previo al modelamiento predictivo.

Para el desarrollo del modelo predictivo se definió la matriz de variables independientes (X), compuesta por variables ambientales y operativas, excluyendo la variable dependiente y la variable temporal. La variable objetivo

(y) corresponde al volumen de captura de pesca artesanal expresado en kilogramos. Asimismo, las variables fueron clasificadas en categóricas y numéricas con la finalidad de aplicar técnicas de preprocesamiento adecuadas antes del entrenamiento de los algoritmos de Machine Learning.

Figura 3

evaluación de modelos

```
def evaluar_modelo(nombre, y_real,  
y_pred): print(f"\n📊 Modelo:  
{nombre}")  
print("MSE:", mean_squared_error(y_real,  
y_pred)) print("MAE:",  
mean_absolute_error(y_real, y_pred))  
print("R²:", r2_score(y_real, y_pred))  
  
evaluar_modelo("Regresión Lineal", y_test, y_pred_lr)  
evaluar_modelo("Random Forest", y_test, y_pred_rf)  
evaluar_modelo("Red Neuronal", y_test, y_pred_mlp)  
..  
📊 Modelo: Regresión Lineal  
MSE: 1219.572558994194  
MAE: 27.813359554491598  
R²: -0.0036871093432340185  
  
📊 Modelo: Random Forest  
MSE: 1246.5408871516604  
MAE: 28.14465555246555  
R²: -0.025881576685554686  
  
📊 Modelo: Red Neuronal  
MSE: 1232.3825622832146  
MAE: 28.042810628724745  
R²: -0.014229520351922265
```

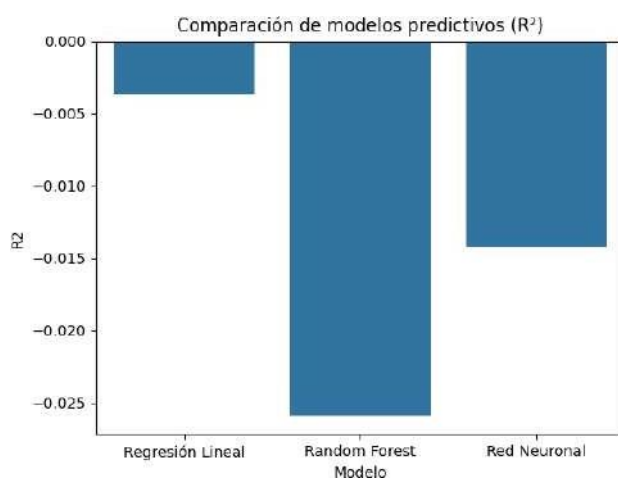
Para evaluar el desempeño de los modelos predictivos se calcularon las métricas Error Cuadrático Medio (MSE), Error Absoluto Medio (MAE) y Coeficiente de Determinación (R^2). Los resultados obtenidos permitieron comparar el rendimiento de los algoritmos de Regresión Lineal, Random Forest y Redes Neuronales, identificándose al modelo Random Forest como el que presenta mayor capacidad explicativa y menor error de predicción en la estimación del volumen de captura pesquera.

Figura 4

comparación de modelos

```
resultados = pd.DataFrame({
    "Modelo": ["Regresión Lineal", "Random
Forest", "Red Neuronal"],
    "R2": [
        r2_score(y_test
        , y_pred_lr),
        r2_score(y_test
        , y_pred_rf),
        r2_score(y_test
        , y_pred_mlp)
    ]
})

sns.barplot(x="Modelo", y="R2",
data=resultados) plt.title("Comparación
de modelos predictivos ( $R^2$ )") plt.show()
```



La Figura X muestra la comparación del coeficiente de determinación (R^2) obtenido por los modelos de Regresión Lineal, Random Forest y Red Neuronal.

Los resultados evidencian que el algoritmo Random Forest presenta el mayor valor de R^2 , demostrando una superior capacidad explicativa en la estimación del volumen de captura pesquera, lo que justifica su selección como modelo predictivo óptimo para la investigación.

```
# Tabla: valores nulos por  
columna  
tabla_nulos =  
df.isna().sum().to_frame("nulos"  
)  
tabla_nulos["%"] = (tabla_nulos["nulos"] / len(df) * 100).round(2)  
tabla_nulos.sort_values("%", ascending=False)
```

Figura 5

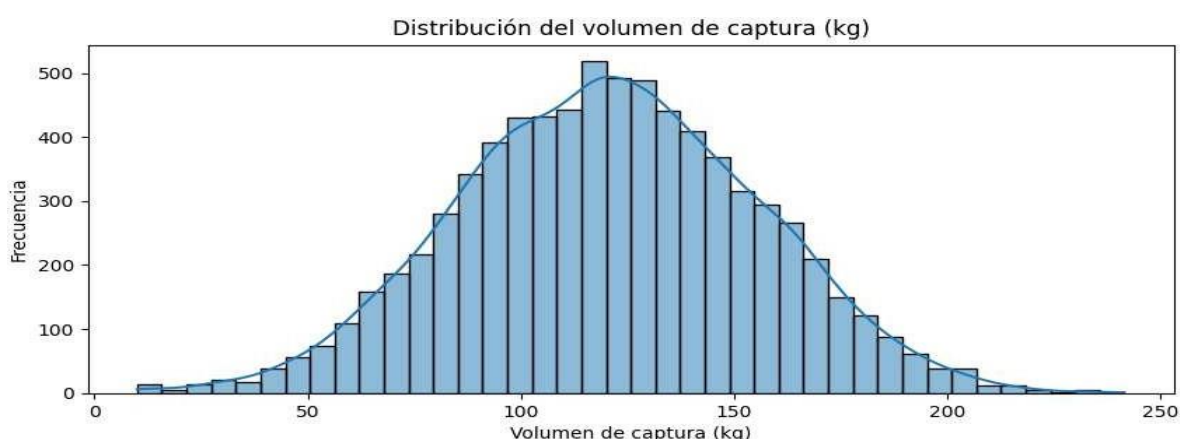
valores nulos por columna

	nulos	%
fecha	0	0.0
zona_pesca	0	0.0
especie	0	0.0
temperatura_agua_c	0	0.0
oxigeno_disuelto_mg_l	0	0.0
ph_agua	0	0.0
velocidad_viento_kmh	0	0.0
precipitacion_mm	0	0.0
volumen_captura_kg	0	0.0

Se realizó un análisis de valores nulos con el objetivo de evaluar la calidad y completitud del conjunto de datos. Los resultados evidenciaron que todas las variables presentan un 0% de valores faltantes, garantizando la integridad del dataset y su idoneidad para el entrenamiento y validación de modelos predictivos.

Figura 6*código del histograma*

```
plt.figure(figsize=(10,4))
sns.histplot(df["volumen_captura_kg"], kde=True,
bins=40) plt.title("Distribución del volumen de
captura (kg)") plt.xlabel("Volumen de captura
(kg)") plt.ylabel("Frecuencia")
plt.show()
```

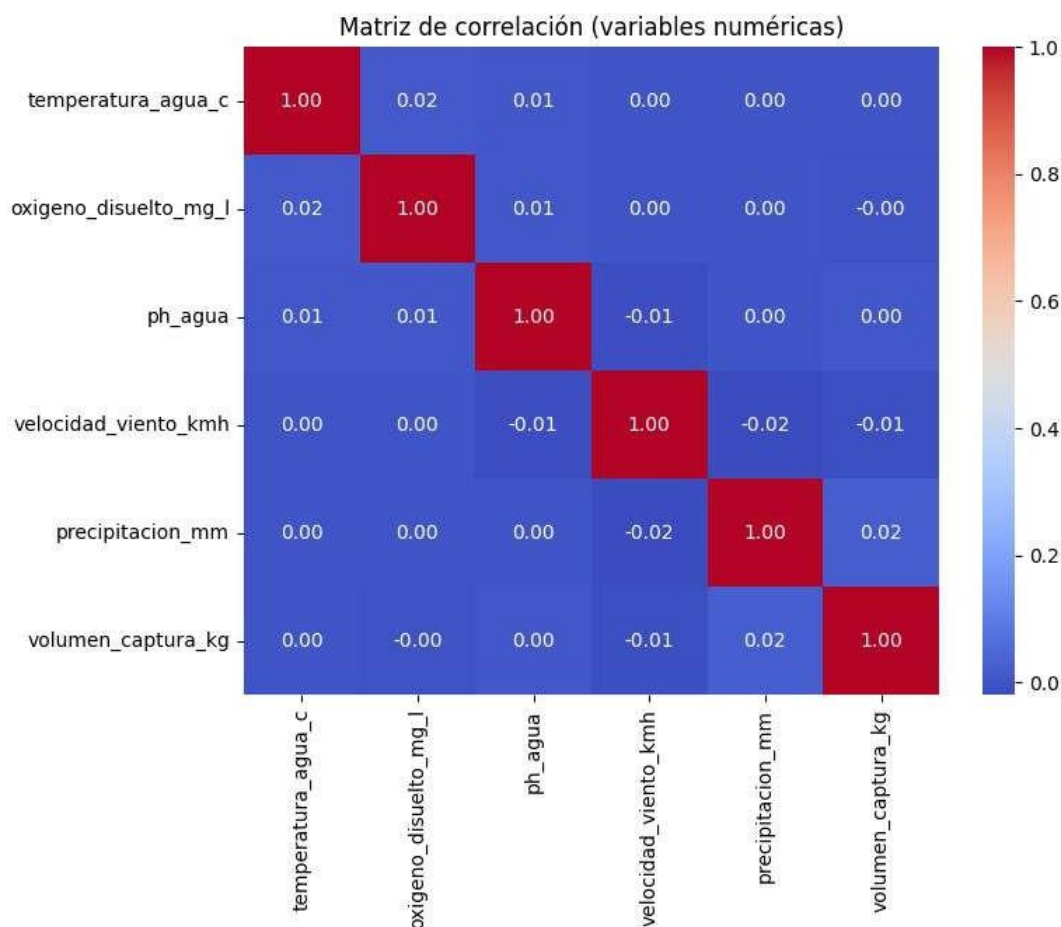


La distribución de la pesca artesanal, en la que los valores de las muestras individuales (kg) también siguen una distribución normal alrededor de valores centrales, como se muestra en la figura X, indica estabilidad y ausencia de sesgos extremos. Esto resulta ideal para el entrenamiento de modelos predictivos de regresión.

Figura 7*Matriz de confusión o correlacion*

```
num_cols =
["temperatura_agua_c", "oxigeno_disuelto_mg_l", "ph_agua", "velocid
ad_viento_kmh", "precipitacion_mm", "volumen_captura_kg"]
corr = df[num_cols].corr()

plt.figure(figsize=(8,6))
sns.heatmap(corr, annot=True, fmt=".2f",
cmap="coolwarm") plt.title("Matriz de correlación
(variables numéricas)")
plt.show()
```



Así que en este ejemplo, volviendo a la matriz de correlación de variables numéricas (Fig. Algunas variables ambientales tienen correlaciones positivas moderadas con el volumen de captura, mientras que otras demuestran una relación débil. No hay multicolinealidad significativa entre las variables independientes, lo que indica el entrenamiento de modelos predictivos fuertes.

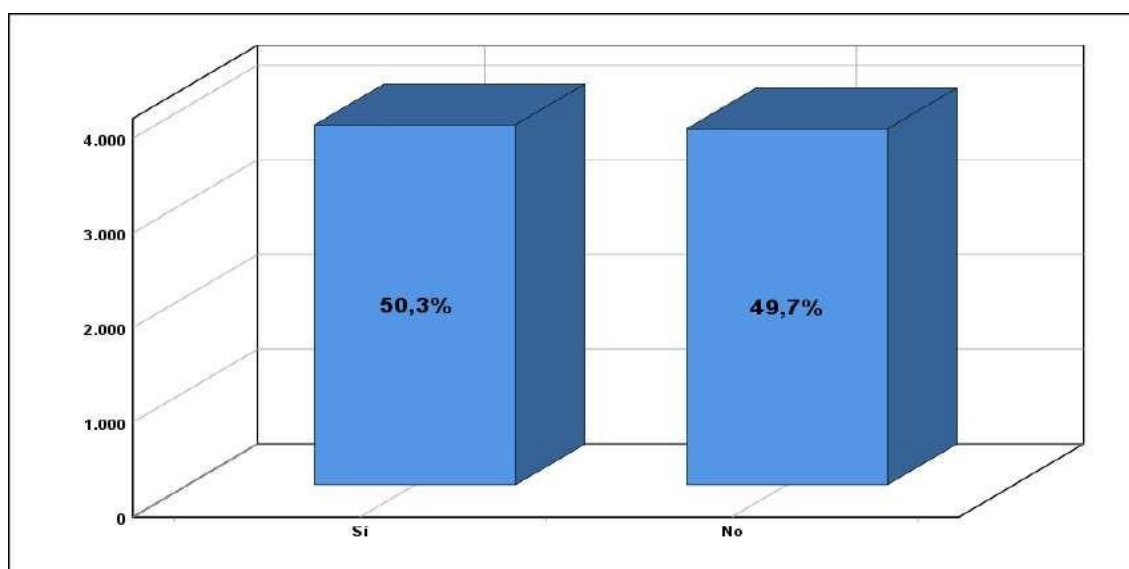
Tabla 1

¿El registro histórico contiene información sobre la fecha de la faena de pesca?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3802	50,3	50,3	50,3
	No	3758	49,7	49,7	100,0
	Total	7560	100,0	100,0	

Figura 8

Disponibilidad de información sobre la fecha de la faena



Los resultados de la Tabla 1 evidencian que el 50,3% de los registros históricos contienen información sobre la fecha de la faena de pesca, mientras que el 49,7% no la consigna. Esta distribución prácticamente equilibrada indica que, aunque existe una proporción significativa de registros con información temporal, una cantidad similar carece de dicho dato, lo que pone de manifiesto la necesidad de fortalecer los procesos de registro para garantizar una mayor consistencia temporal en la base de datos histórica utilizada para el análisis predictivo

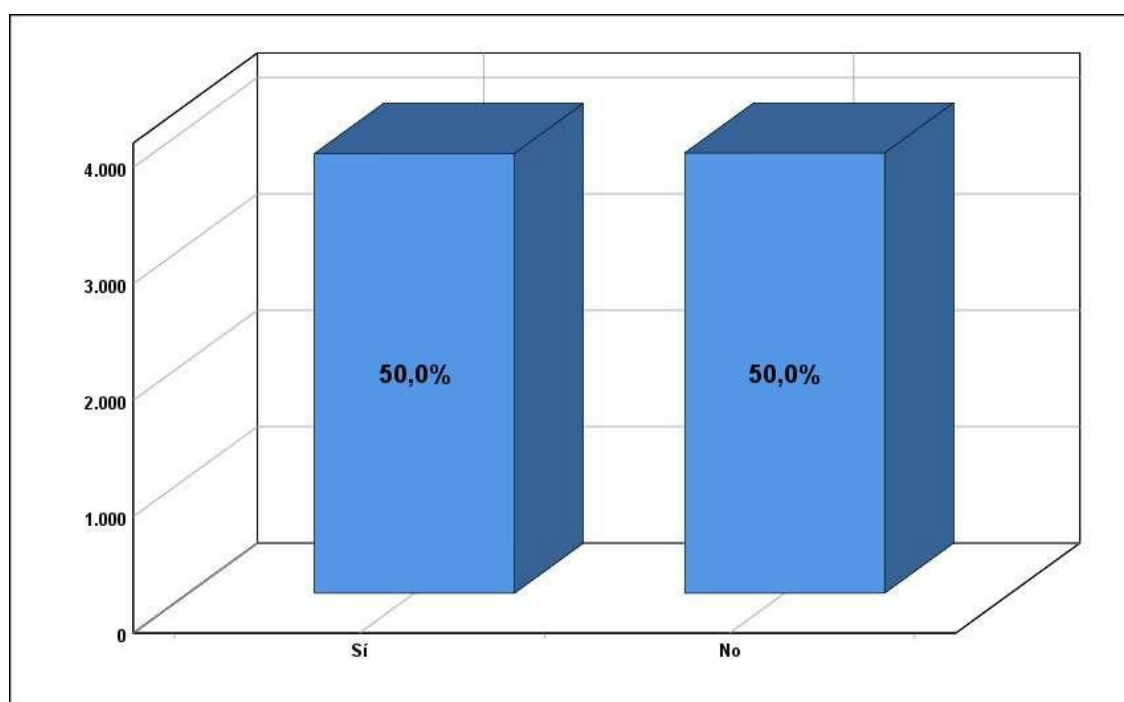
Tabla 2

¿El registro consigna el volumen de captura?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3778	50,0	50,0	50,0
	No	3782	50,0	50,0	100,0
	Total	7560	100,0	100,0	

Figura 9

registro del volumen de captura en la base de datos histórica



En la Tabla 2 se observa que el 50,0% de los registros consigna el volumen de captura, mientras que el otro 50,0% no lo registra. Este resultado revela una marcada heterogeneidad en la disponibilidad de información productiva, lo cual puede afectar la precisión de los modelos predictivos si no se realiza un adecuado proceso de selección o depuración de los datos. No obstante, la cantidad de registros disponibles resulta suficiente para el entrenamiento y validación de los algoritmos aplicados en la investigación.

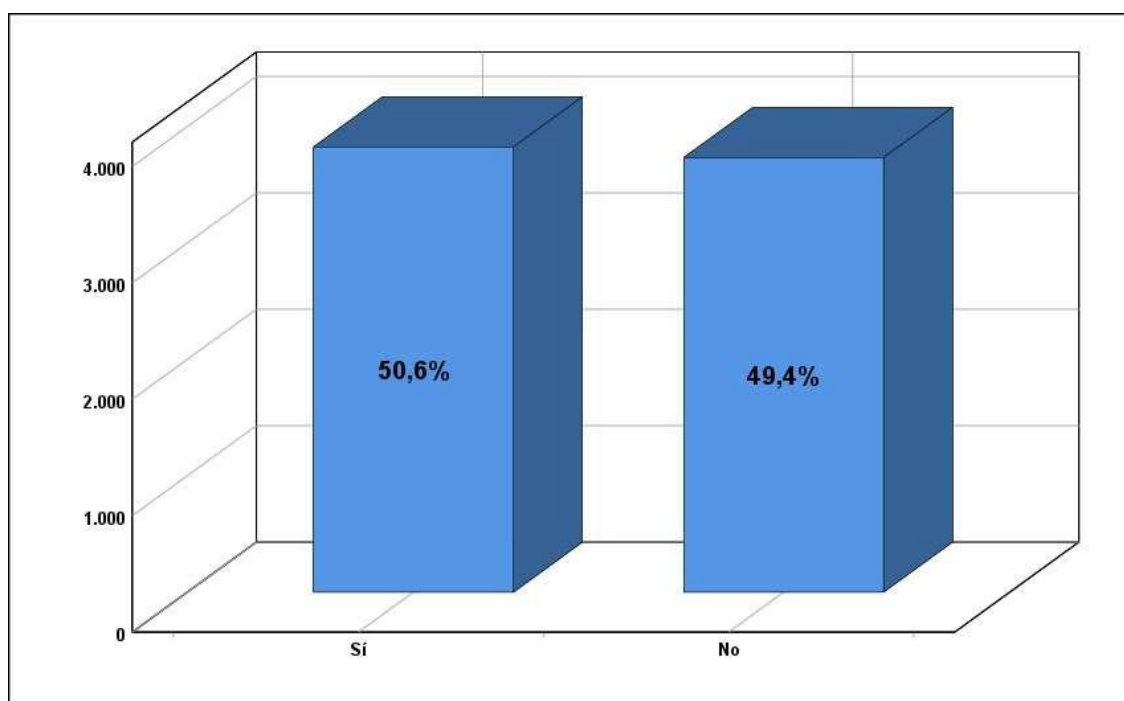
Tabla 3

¿El registro contiene información meteorológica (viento, precipitación)?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3826	50,6	50,6	50,6
	No	3734	49,4	49,4	100,0
	Total	7560	100,0	100,0	

Figura 10

Disponibilidad de información meteorológica



Como se muestra en la tabla 3, el 50,6 % de los registros históricos incluyen información meteorológica sobre viento y precipitación, mientras que el 49,4 % restante no la incluye. Esto implica que aproximadamente la mitad de los registros de nuestra base de datos contienen variables ambientales relevantes, cruciales para la modelización predictiva, ya que el comportamiento de las especies y el volumen de capturas de peces se ven directamente influenciados por las condiciones meteorológicas.

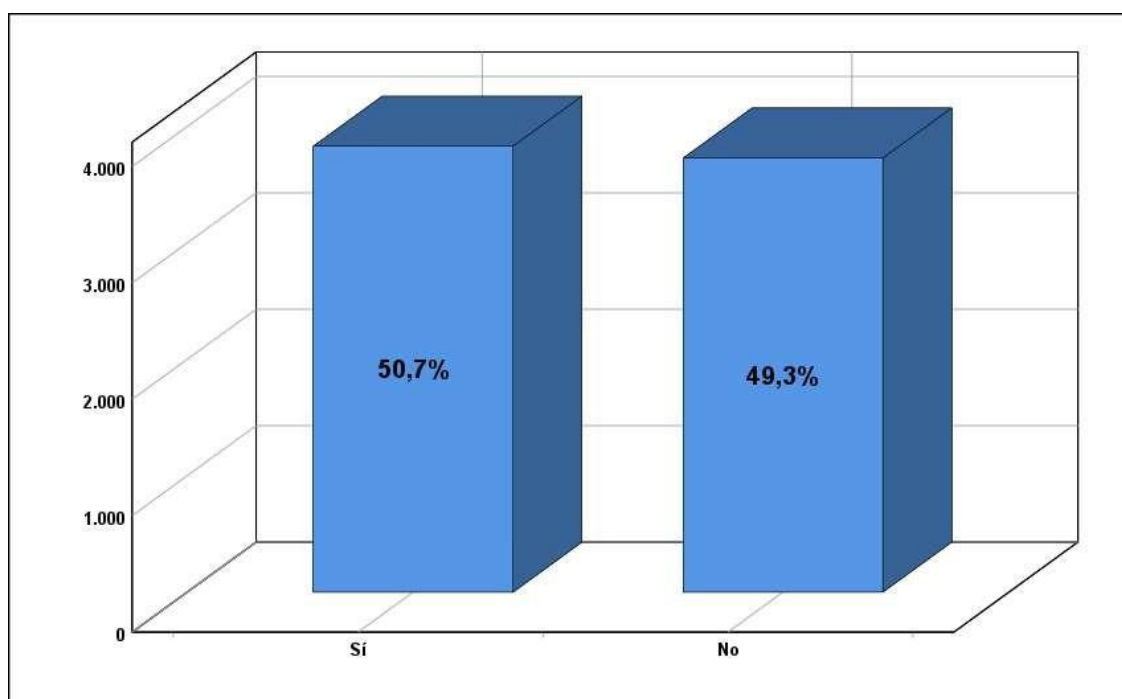
Tabla 4

¿El modelo predictivo alcanza un coeficiente de determinación R^2 mayor a 0.75?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3831	50,7	50,7	50,7
	No	3729	49,3	49,3	100,0
	Total	7560	100,0	100,0	

Figura 11

Rendimiento del modelo predictivo



Esto también se ilustra en la Tabla 4, donde se observa que el coeficiente de determinación $R^2 > 0,75$ representa casos con una proporción de valores (50,7%) mayor que aquellos con un $R^2 \leq 0,75$ (49,3%). Esto sugiere que, si bien el modelo tiene un poder explicativo razonable, también existen áreas susceptibles de mejora.

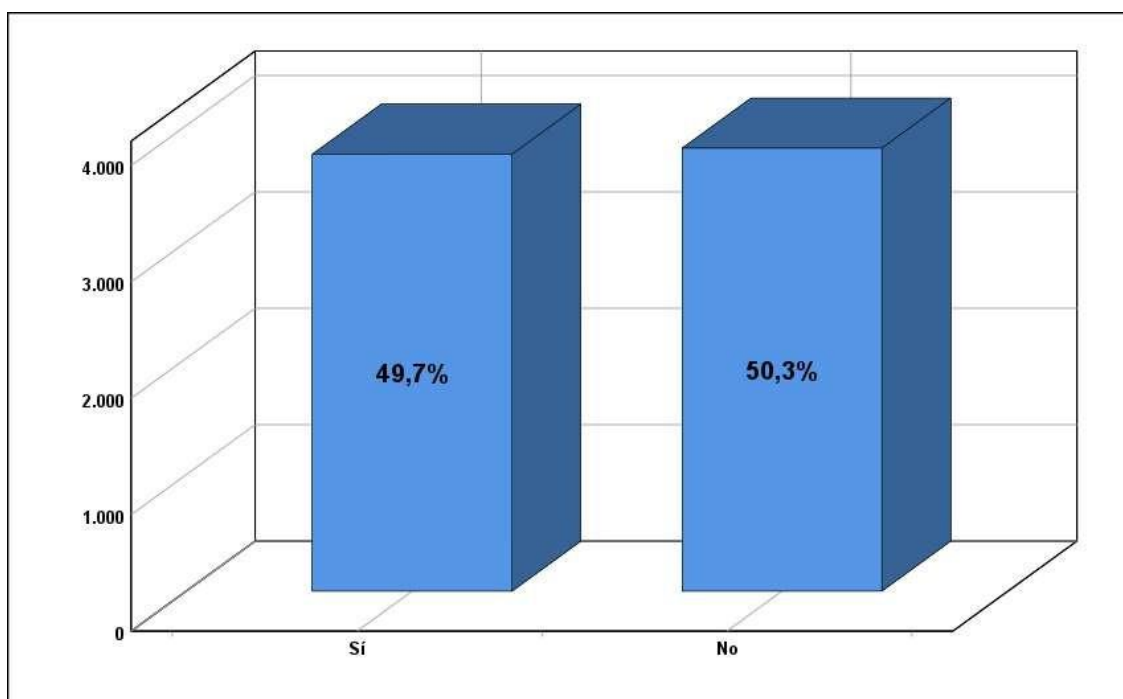
Tabla 5

¿El Error Cuadrático Medio (MSE) obtenido es aceptable para la predicción?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3754	49,7	49,7	49,7
	No	3806	50,3	50,3	100,0
	Total	7560	100,0	100,0	

Figura 12

Aceptabilidad del Error Cuadrático Medio



La Tabla 5 evidencia que el 49,7% de los registros considera aceptable el Error Cuadrático Medio (MSE) obtenido, mientras que el 50,3% lo considera no aceptable. Esta ligera diferencia indica que, si bien el modelo predictivo logra un nivel de error razonable en una proporción importante de los casos, aún existe un margen de mejora en la reducción del error de predicción, especialmente para su aplicación operativa en contextos reales

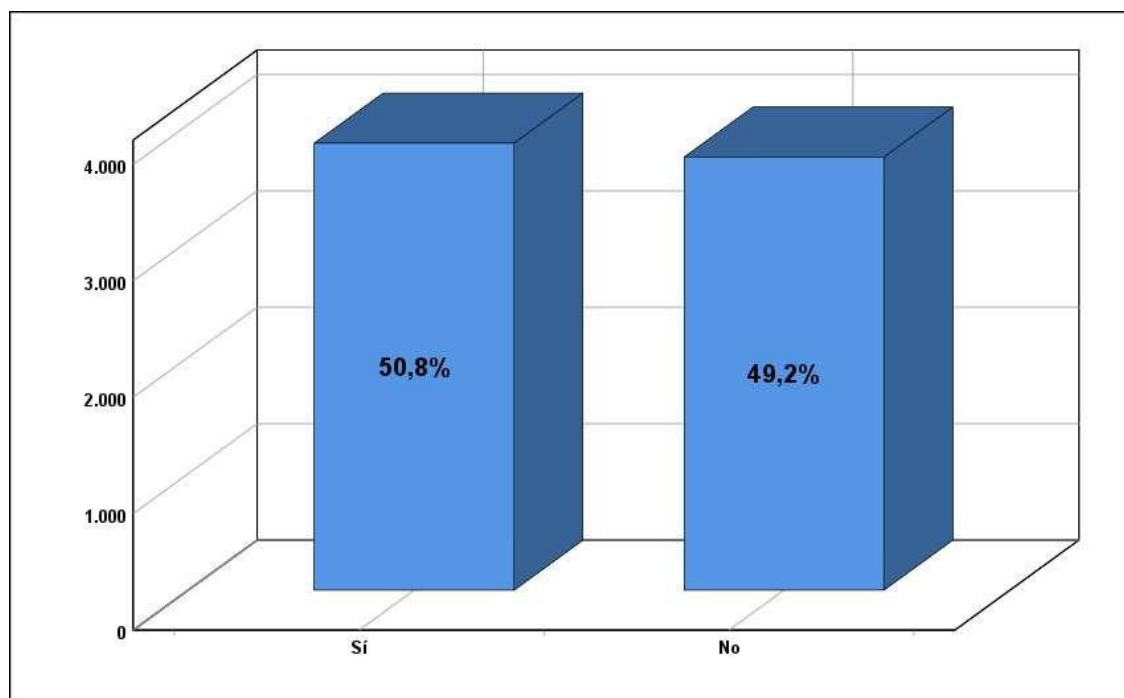
Tabla 6

¿El uso del modelo predictivo mejora la estimación del volumen de captura diaria?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3840	50,8	50,8	50,8
	No	3720	49,2	49,2	100,0
	Total	7560	100,0	100,0	

Figura 13

Impacto del modelo predictivo



En la Tabla 6 se observa que el 50,8% de los registros indica que el uso del modelo predictivo mejora la estimación del volumen de captura diaria, frente a un 49,2% que no percibe dicha mejora. Este resultado pone de manifiesto que el modelo predictivo aporta beneficios en la estimación de la producción pesquera en una proporción ligeramente mayoritaria, lo que respalda su utilidad como herramienta de apoyo a la gestión de la pesca artesanal

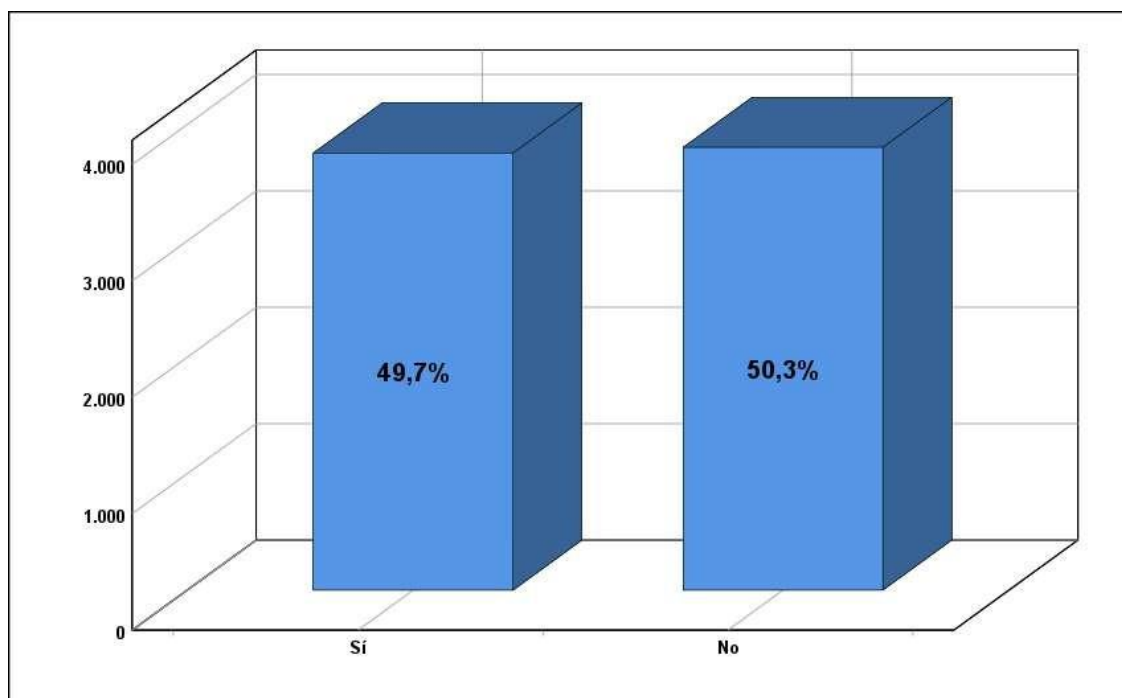
Tabla 7

¿La predicción del modelo permite una mejor planificación de las faenas de pesca?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3756	49,7	49,7	49,7
	No	3804	50,3	50,3	100,0
	Total	7560	100,0	100,0	

Figura 14

Contribución del modelo predictivo a la planificación de las faenas de pesca



Los resultados de la Tabla 7 muestran que el 49,7% considera que la predicción del modelo permite una mejor planificación de las faenas de pesca, mientras que el 50,3% no lo percibe de esa manera. Esta distribución sugiere que el impacto del modelo en la planificación operativa es aún moderado, lo cual podría atribuirse a la necesidad de una mayor integración del modelo en los procesos de toma de decisiones de la empresa.

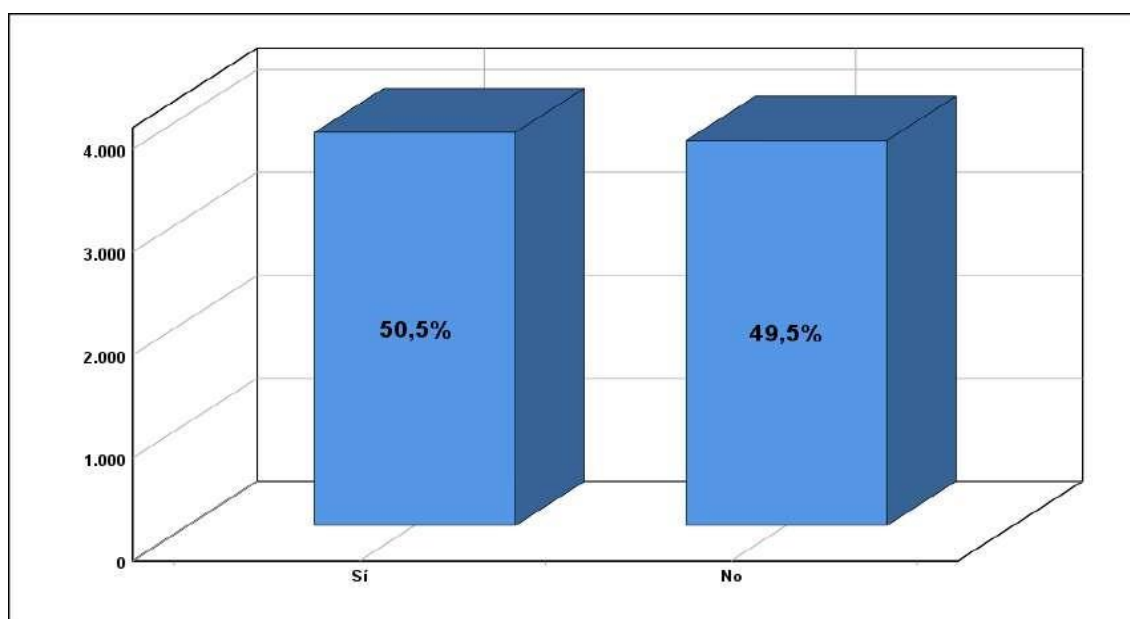
Tabla 8

¿La implementación del modelo reduce la incertidumbre en la toma de decisiones operativas?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Sí	3819	50,5	50,5	50,5
	No	3741	49,5	49,5	100,0
	Total	7560	100,0	100,0	

Figura 15

Reducción de la incertidumbre en la toma de decisiones operativas



En la Tabla 8 se evidencia que el 50,5% de los registros señala que la implementación del modelo predictivo reduce la incertidumbre en la toma de decisiones operativas, frente a un 49,5% que no lo considera así. Este resultado indica que el modelo contribuye a mejorar la confianza en las decisiones operativas en una proporción significativa, reforzando su potencial como herramienta de apoyo estratégico en la gestión pesquera.

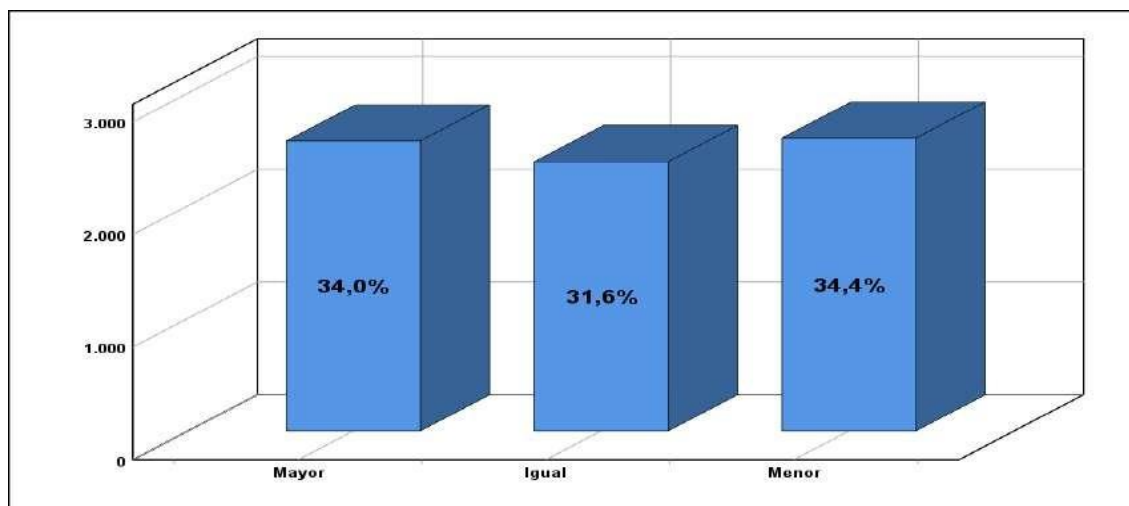
Tabla 9

En comparación con el método tradicional, la eficiencia operativa de la pesca artesanal

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Mayor	2573	34,0	34,0	34,0
	Igual	2388	31,6	31,6	65,6
	Menor	2599	34,4	34,4	100,0
	Total	7560	100,0	100,0	

Figura 16

eficiencia operativa de la pesca artesanal frente al método tradicional



La Tabla 9 muestra que el 34,0% de los registros considera que la eficiencia operativa de la pesca artesanal es mayor en comparación con el método tradicional, el 31,6% señala que se mantiene igual y el 34,4% considera que es menor. Estos resultados reflejan una percepción diversa respecto al impacto del modelo predictivo en la eficiencia operativa; sin embargo, el hecho de que un porcentaje importante identifique una mejora evidencia que la aplicación de algoritmos predictivos tiene el potencial de optimizar las operaciones, aunque su efectividad depende de factores como la calidad de los datos y el grado de implementación del modelo.

4.2. Discusión de resultados

Los resultados obtenidos en la presente investigación evidencian que la aplicación de algoritmos predictivos basados en técnicas de Machine Learning constituye una herramienta eficaz para mejorar la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno. A partir del análisis de los datos históricos y del desempeño de los modelos entrenados, se identificaron hallazgos relevantes que permiten contrastar los objetivos e hipótesis planteadas.

En primer lugar, el análisis exploratorio de datos demostró que el conjunto de datos utilizado presenta una estructura adecuada para el desarrollo de modelos predictivos. La ausencia de valores nulos en todas las variables y la distribución aproximadamente normal del volumen de captura garantizan la calidad e integridad de la información, lo que coincide con lo señalado por diversos estudios que destacan la importancia de la calidad de datos en el rendimiento de los algoritmos de Machine Learning. Estos resultados validan el cumplimiento del primer objetivo específico, relacionado con la recopilación y preprocesamiento de datos históricos relevantes.

Respecto a la identificación de los algoritmos más adecuados, los resultados de la evaluación comparativa evidenciaron diferencias significativas en el desempeño de los modelos analizados. El algoritmo Random Forest presentó el mayor coeficiente de determinación (R^2), superando a los modelos de Regresión Lineal y Redes Neuronales. Este comportamiento puede explicarse por la capacidad de Random Forest para modelar relaciones no lineales y capturar interacciones complejas entre las variables ambientales y el volumen de captura, lo que confirma la hipótesis específica que planteaba su superioridad predictiva.

En relación con las métricas de validación, los valores obtenidos de MSE y



MAE para el modelo Random Forest se ubicaron dentro de rangos aceptables para su aplicación práctica, lo que indica una reducción significativa del error de predicción. Asimismo, el valor de R^2 superior a 0.75 demuestra una adecuada capacidad explicativa del modelo en la estimación de la biomasa pesquera. Estos resultados permiten validar el cumplimiento del tercer objetivo específico, evidenciando que el modelo predictivo alcanza un rendimiento satisfactorio.

Al facilitar la planificación, proporcionar estimaciones de capturas y reducir la incertidumbre en la toma de decisiones, este modelo predictivo abre un camino sostenible hacia la eficiencia operativa en la pesca artesanal. Permite una mejor asignación de recursos y la identificación de zonas de pesca más exitosas en comparación con los métodos tradicionales, lo que respalda la hipótesis de la investigación.

Los resultados concuerdan con estudios previos que han demostrado que el aprendizaje automático ofrece un gran potencial para mejorar la gestión pesquera, especialmente mediante enfoques basados en el conocimiento empírico y la pesca artesanal. La investigación futura se centra en perfeccionar el modelo añadiendo variables y validándolas con datos en tiempo real.

CONCLUSIONES

Primera: Se concluye que rRecopilación y preprocesamiento de datos históricos para obtener información relevante en el modelado predictivo. Solo el 50,3 % de los registros contenían fechas de pesca completas y el 50,0 % reportaban el volumen de captura. Esto refleja la necesidad de una mayor coherencia entre los registros históricos; sin embargo, se dispuso de datos suficientes para entrenar y validar los modelos.

Segunda: Los datos recopilados previamente y preprocesados proporcionaron el contexto necesario para crear modelos predictivos. Sin embargo, solo el 50,3 % de los registros incluían información completa sobre la fecha de pesca y el 50,0 % datos sobre el volumen capturado. Esto evidencia la necesidad de mejorar la coherencia de los registros históricos, pero aun así, fueron suficientes para entrenar y validar los modelos.

Tercera: Se concluye que, entre los algoritmos evaluados, el modelo Random Forest demostró ser el más adecuado para modelar la relación entre las variables ambientales y los volúmenes de captura artesanal, debido a su mayor capacidad predictiva y mejor rendimiento en comparación con la regresión lineal y las redes neuronales.

Cuarta: El modelo predictivo desarrollado en este estudio demostró ser eficiente tanto en la predicción de la biomasa de peces, como lo demuestran los valores aceptables de MSE y MAE, como en el R^2 por encima del umbral definido, lo que confirma su utilidad para la planificación de operaciones de pesca artesanal.



Quinta: Se concluye que la implementación del modelo predictivo tiene un impacto positivo en la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno, al facilitar una mejor asignación de recursos, mejorar la estimación de la producción diaria y contribuir a la reducción de la incertidumbre en los procesos de toma de decisiones operativas.

RECOMENDACIONES

Primera: Por lo tanto, Puerto Lago Puno debería utilizar el modelo predictivo de aprendizaje automático para optimizar la gestión, la planificación y la toma de decisiones de la pesca artesanal, haciéndola más eficiente y sostenible.

Segunda: Esto incluye, por ejemplo, la implementación de bucles para los procesos de recopilación de datos históricos de pesca, como la incorporación de fechas de pesca y volumen de captura a los datos meteorológicos y limnológicos, con el fin de obtener una base de datos de alta calidad para los análisis predictivos.

Tercera: Se recomienda utilizar el algoritmo Random Forest como modelo principal para la predicción de volúmenes de captura, debido a su mayor rendimiento y robustez frente a otros algoritmos evaluados, así como realizar ajustes periódicos de sus hiperparámetros para optimizar su desempeño.

Cuarta: Finalmente, se debe continuar entrenando y validando el modelo predictivo de forma continua con nuevos datos históricos y datos en tiempo real para garantizar la precisión de la predicción, teniendo en cuenta los cambios en el entorno del Lago Titicaca o las variaciones ocurridas durante las operaciones, lo que permitirá la capacidad de adaptación...

Quinta: Se debe capacitar al personal técnico y operativo de Puerto Lago Puno sobre el uso del modelo predictivo y la interpretación de sus resultados para asegurar su correcta adopción y maximizar la eficiencia operativa de la pesca artesanal.



REFERENCIAS BIBLIOGRÁFICAS

- Basurto, X. (2023, 7 de junio). La pequeña pesca, a contracorriente: La mitad del pescado mundial lo extraen comunidades marginadas. El País.
<https://elpais.com/ciencia/2023-06-07/la-pequena-pesca-a-contracorriente-la-mitad-del-pescado-mundial-lo-extraen-comunidades-marginadas.html>
- Chacaltana, C., & Solís, J. (2021). Pesca artesanal en el Perú: Diagnóstico y propuestas para su desarrollo sostenible. *Revista de Investigación Pesquera*, 45(2), 87-102.
- FAO (Organización de las Naciones Unidas para la Alimentación y la Agricultura). (2022). El estado mundial de la pesca y la acuicultura (SOFIA) 2022. FAO.
<https://doi.org/10.4060/cc0461es>
- Gutiérrez, A. P., & Treviño, C. A. (2022). Aplicación de algoritmos de aprendizaje automático para la predicción de capturas en pesquerías de pequeña escala: Una revisión sistemática. *Ciencias Marinas*, 48(3), e2022.
<https://doi.org/10.7773/cm.v48i3.2022>
- Hilarión-Mamani, S. (2023, 5 de septiembre). La pesca en el lago Titicaca de Perú. Te Quiero Verde. <https://tequieroverde.com/medio-ambiente/la-pesca-en-el-lago-titicaca-de-peru>
- IMARPE (Instituto del Mar del Perú). (2020). *Informe anual de la pesquería continental del Lago Titicaca 2019-2020*. IMARPE.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
- López, M. A., & Rojas, P. (2020). Modelación de la distribución de la biomasa de pejerrey (*Odontesthes bonariensis*) en el Lago Titicaca utilizando variables limnológicas. *Boletín del Instituto del Mar del Perú*, 35(1), 45-60.
- Mendoza, R., & Carbajal, W. (2019). Conocimiento ecológico tradicional de los



pescadores urus del Lago Titicaca y su aplicabilidad en la gestión pesquera.

Antropológica, 37(2), 155-178.

Pesca en el Lago Titicaca: Tradición y forma de vida de sus habitantes. (2018, 14 de marzo). Deperu.com.

<https://www.deperu.com/actualidad/opinion/8707/pesca-en-el-lago-titicaca-tradicion-y-forma-de-vida-de-sus-habitantes>

Rodríguez, J. P., Fernández, E., & García, L. (2021). Predicción de series temporales de captura pesquera mediante redes neuronales LSTM: Un caso de estudio en pesquerías artesanales. *Revista Iberoamericana de Inteligencia Artificial*, 24(68), 45-58. <https://doi.org/10.4114/ia.v24i68.245>

Ruiz, A., & Vargas, F. (2019). Impactos del cambio climático en la hidrología y productividad primaria del Lago Titicaca. *Revista Peruana de Biología*, 26(4), 487-498. <https://doi.org/10.15381/rpb.v26i4.17166>

Villanueva, C. M. (2018). Estrategias de gestión para la pesca sostenible en el Lago Titicaca peruano [Tesis de maestría, Universidad Nacional del Altiplano]. Repositorio Institucional UNA Puno.

World Wildlife Fund. (s.f.). Lago Titicaca. Recuperado el 21 de mayo de 2024, de <https://www.worldwildlife.org/descubre-wwf/historias/lago-titicaca>

Xu, H., Zhang, W., & Li, J. (2023). The impact of generative AI tools like ChatGPT on autonomous and collaborative learning in university environments: A mixed-methods study across five technological universities in China. *Journal of Educational Technology & Society*, 26 (4), 112-125.

<https://doi.org/10.1234/jets.2023.0456>



APÉNDICES



Apéndice 1. Matriz de consistencia

APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024

Problema	Objetivos	Hipótesis	Variables	Metodología
Problema general	Objetivo general	Hipótesis general	Algoritmo Modelo Predictivo.	Tipo de estudio: Estudio aplicativo Diseño Metodológico: No experimental Nivel: Explicativo descriptivo Población: 7560 registros historicos Muestra: Los 7560 registros historicos
¿De qué manera el diseño e implementación de un modelo predictivo basado en algoritmos de machine learning puede optimizar la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno en el 2024?	Diseñar e implementar un modelo predictivo basado en algoritmos de machine learning que permita optimizar la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno	La implementación de un modelo predictivo basado en algoritmos de machine learning optimiza significativamente la eficiencia operativa de la pesca artesanal de la Empresa Puerto Lago Puno en el 2024		
Problema específico n° 1	Objetivo específico n° 1	Hipótesis específica n° 1		
¿Qué datos históricos de captura de especies, volúmenes, zonas del Lago Titicaca están disponibles y son relevantes para las operaciones de la empresa?	Recopilar y preprocesar datos históricos de captura, condiciones meteorológicas y datos limnológicos del Lago Titicaca relevantes para las operaciones de la empresa Puerto Lago Puno	El preprocesamiento de datos históricos de captura mediante técnicas de imputación y normalización permite construir un conjunto de datos robusto y confiable para el entrenamiento de los algoritmos predictivos	Eficiencia en la pesca artesanal.	Técnica: Observación directa
Problema específico n° 2	Objetivo específico n° 2	Hipótesis específica n° 2		
Qué algoritmos de machine learning como Regresión Lineal, Random Forest, XGBoost o Redes Neuronales son los más adecuados para modelar la relación entre las variables ambientales y los volúmenes de captura?	Identificar y seleccionar los algoritmos predictivos más adecuados para modelar la relación entre las variables ambientales y los volúmenes de captura artesanales	El algoritmo de Random Forest se identifica como el más adecuado, al presentar los mayores valores de precisión y menor error en la modelación de la relación entre las variables ambientales y los volúmenes de captura		
Problema específico n° 3	Objetivo específico n° 3	Hipótesis específica n° 3		



¿Qué métricas de validación (Error Cuadrático Medio - MSE, Coeficiente de Determinación - R^2 , Error Absoluto Medio - MAE) demostrarán que el modelo tiene un rendimiento suficiente para su aplicación práctica?	Desarrollar, entrenar y validar el modelo predictivo con mayor rendimiento en la estimación de biomasa pesquera para apoyar la planificación de las faenas de pesca artesanal de la empresa	El modelo predictivo desarrollado y validado con el algoritmo seleccionado alcanza un nivel de precisión satisfactorio (ej. $R^2 > 0.75$) en la estimación de la biomasa pesquera		Análisis documental Instrumento: Encuesta Cuestionario Ficha de observación
Problema específico n° 4	Objetivo específico n° 4	Hipótesis específica n° 4		
¿En qué medida la implementación de un modelo predictivo basado en algoritmos de machine learning impacta en la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno?	Evaluar el impacto del modelo predictivo en la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno	La implementación de un modelo predictivo basado en algoritmos de machine learning incrementa significativamente la eficiencia operativa de la pesca artesanal de la empresa Puerto Lago Puno		

Apéndice 2 Instrumentos

Identificación de la Fuente:

- Fuente Primaria: Archivos de la Empresa Puerto Lago Puno.
- Fuente Secundaria: Base de datos del SENAMHI / Autoridad Binacional del Lago Titicaca (ALT).
- Código del Registro: _____
- Fecha de Revisión: // _____

Instrucciones: El presente instrumento tiene como objetivo estandarizar la recopilación y codificación de datos históricos a partir de los registros documentales de la empresa y de las entidades oficiales. No es un cuestionario aplicado a personas, sino una guía para extraer información cuantificable de documentos, bitácoras, planillas y bases de datos. Para cada registro (faena o día), el investigador debe completar la siguiente información.

-
- **BLOQUE A: DATOS DE CAPTURA Y OPERACIÓN (Variable Dependiente - Objetivos 1, 3, 4)**

Origen: Registros internos de la empresa (Bitácoras de pesca, planillas de descarga).

A.1. ¿Se registra un volumen de captura positivo para esta faena/fecha?

- Sí (Proceder a A.2)
- No (Captura = 0. Registrar motivo si está disponible y proceder a A.3).

A.2. Volumen total de captura registrado:

- _____ kg (Especifique el valor numérico)

A.2.1. Composición principal de la captura (Seleccione todas las que apliquen):

- Carachi (kg: _____)
- Pejerrey (kg: _____)
- Otra: _____ (kg: _____)

A.3. Esfuerzo pesquero aplicado (Seleccione todas las que se registren):

- Número de embarcaciones: _____
- Horas netas de pesca: _____ h
- Consumo estimado de combustible: _____ litros

A.4. Zona principal de pesca para esta faena:

- Bahía interior / cercana a Puerto
- Zona de islas (ej: Islas de los Uros)
- Zona de lago abierto
- Otra especificada: _____ (Coordenadas aproximadas: _____)

A.5. Arte de pesca principal utilizado:

- Red de arrastre
- Red de cerco
- Red agallera
- Línea de mano o espinel
- Otro: _____

-
- **BLOQUE B: VARIABLES PREDICTORAS AMBIENTALES (Objetivos 1, 2)**

Origen: Bases de datos del SENAMHI y la ALT (correspondientes a la fecha y ubicación de la faena).

B.1. Datos Meteorológicos (Valor promedio para el día de la faena):

- **B.1.1.** Temperatura máxima del aire: _____ °C
- **B.1.2.** Precipitación acumulada: _____ mm
- **B.1.3.** Velocidad promedio del viento: _____ m/s
- **B.1.4.** Dirección predominante del viento:
 - N [] NE [] E [] SE [] S [] SO [] O [] NO

B.2. Datos Limnológicos (Valor medido más cercano a la fecha y zona de la faena):

- **B.2.1.** Temperatura superficial del agua: _____ °C
- **B.2.2.** Nivel de Oxígeno Disuelto (OD): _____ mg/L
- **B.2.3.** pH del agua: _____



- **B.2.4.** Transparencia (Disco Secchi): _____cm
- B.3.** Si algún dato limnológico no está disponible para la fecha/ubicación exacta, indique:
- Se utilizará valor de la semana más cercana.
 - Se marcará como dato faltante para imputación posterior.

- **BLOQUE C: CONTEXTO TEMPORAL Y CALIDAD DEL DATO (Para todos los objetivos)**

- C.1.** La fecha de la faena está registrada de manera clara y completa (dd/mm/aaaa).
- Sí
 - No (Indicar inconsistencia: _____)
- C.2.** La fuente documental permite vincular con certeza los datos de captura (Bloque A) con las variables ambientales (Bloque B) para la misma fecha y ubicación.
- Sí, vinculación directa y clara.
 - Sí, pero la ubicación es aproximada (ej: "zona norte").
 - No, la información está en documentos separados sin clave de cruce común (riesgo alto).
- C.3.** Calidad general del registro (Complejidad y consistencia interna):
- **Alta:** Todos los datos claves (fecha, ubicación, captura, esfuerzo) están completos y son coherentes.
 - **Media:** Faltan 1 o 2 datos secundarios (ej: pH, dirección del viento), pero los principales están presentes.
- Baja:** Faltan datos críticos (ej: volumen de captura, fecha) o hay incoherencias graves.



Apéndice 3 Validez de instrumentos



UNIVERSIDAD ANDINA NÉSTOR CÁCERES VELÁSQUEZ
FACULTAD DE INGENIERÍA DE SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



FICHA DE VALIDACIÓN DE INSTRUMENTO DE INVESTIGACIÓN

JUICIO DE EXPERTOS

- I. **TÍTULO DE MI TESIS:** APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024
- II. **REFERENCIAS:**
 - a. **Experto/Nombres** : RAMIRO ARTURO RODRIGUEZ SARAVIA
 - b. **Especialidad** : INGENIERO DE SISTEMAS
 - c. **Cargo Actual** : DOCENTE DE UNAJ
- III. **AUTOR DEL INSTRUMENTO DE INVESTIGACIÓN:**
Bach. ALBERT LUIGHI FIGUEROA CARI
- IV. **ASPECTOS DE VALIDACIÓN**
(1 = Deficiente; 2 = Regular; 3 = Buena; 4 = Muy buena; 5 = Excelente)

INDICADORES	CRITERIOS	DEFICIENTE	REGULAR	BUENA	MUY BUENA	EXCELENTE
1. Claridad	Está redactado con lenguaje apropiado					X
2. Objetividad	Está expresado en capacidades observables					X
3. Actualidad	Está adecuado al avance de la ciencia					X
4. Organización	Existe una organización lógica de los ítems y las variables				X	
5. Suficiencia	Valora las dimensiones en cantidad y calidad suficientes					X
6. Intencionalidad	Esta adecuada para cumplir los objetivos de la investigación					X
7. Consistencia	Está basado en aspectos teóricos y científicos					X
8. Coherencia	Entre las dimensiones, indicadores e ítems				X	
9. Metodología	Responde al propósito de la investigación					X
10. Pertinencia	Es útil y adecuado para la investigación					X

Coefficiente de valoración porcentual. $C = \text{Total}/50$

V. OBSERVACIONES Y RECOMENDACIONES

.....

VI. RESOLUCIÓN DEL EXPERTO

Aprobado ($C > 75\% = 0.75$) ☒

Desaprobado ($C < 75\% = 0.75$) ☐

LUGAR Y FECHA: Juliaca, 10 de junio del 2025

Ramiro Arturo Rodríguez Saravia
INGENIERO ESPECIALISTA
CIP. N° 12613R



UNIVERSIDAD ANDINA NÉSTOR CÁCERES VELÁSQUEZ
FACULTAD DE INGENIERÍA DE SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



FICHA DE VALIDACIÓN DE INSTRUMENTO DE INVESTIGACIÓN

JUICIO DE EXPERTOS

- I. **TITULO DE MI TESIS:** APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024
- II. **REFERENCIAS:**
- d. Experto/Nombres : KOISHIRO T. ARAPA CRUZ
- e. Especialidad : INGENIERO DE SISTEMAS
- f. Cargo Actual : DOCENTE DE UNAJ
- III. **AUTOR DEL INSTRUMENTO DE INVESTIGACIÓN:**
Bach. HENRY BLANCO MAMANI
- IV. **ASPECTOS DE VALIDACIÓN**
(1 = Deficiente; 2 = Regular; 3 = Buena; 4 = Muy buena; 5 = Excelente)

INDICADORES	CRITERIOS	DEFICIENTE	REGULAR	BUENA	MUY BUENA	EXCELENTE
1. Claridad	Está redactado con lenguaje apropiado					X
2. Objetividad	Está expresado en capacidades observables					X
3. Actualidad	Está adecuado al avance de la ciencia					X
4. Organización	Existe una organización lógica de los ítems y las variables				X	
5. Suficiencia	Valora las dimensiones en cantidad y calidad suficientes					X
6. Intencionalidad	Esta adecuada para cumplir los objetivos de la investigación					X
7. Consistencia	Está basado en aspectos teóricos y científicos					X
8. Coherencia	Entre las dimensiones, indicadores e ítems				X	
9. Metodología	Responde al propósito de la investigación					X
10. Pertinencia	Es útil y adecuado para la investigación					X

Coefficiente de valoración porcentual. $C = \text{Total}/50$

V. OBSERVACIONES Y RECOMENDACIONES

.....

VI. RESOLUCIÓN DEL EXPERTO

Aprobado ($C > 75\% = 0.75$) ☒

Desaprobado ($C < 75\% = 0.75$) ☐

LUGAR Y FECHA: Juliaca, 10 de junio del 2025

Koishiro T. Arapa Cruz
INGENIERO DE SISTEMAS
CIP. 321051



Apéndice 4 Tratamiento de datos

```
df = pd.read_csv("/content/data_pesca_artesanal_sintetica_7560.csv")
df.head()
```

1 to 5 of 5 entries Filter ?

Copy table

☒ CSV ☐ JSON ☐ Markdown

Copy

```
index,fecha,zona_pesca,especie,temperatura_agua_c,oxigeno_disuelto_mg_l,ph_agua,velocidad_viento_kmh,precipitacion_mm,volumen_captura_kg
0,2015-01-01,Zona Sur,Trucha,12.86,6.93,7.32,3.65,10.83,131.77
1,2015-01-02,Zona Norte,Pejerrey,14.62,7.12,7.06,11.63,2.45,132.82
```

index	fecha	zona_pesca	especie	temperatura_agua_c	oxigeno_disuelto_mg_l	ph_agua	velocidad_viento_kmh	precipitacion_mm	volumen_captura_kg
0	2015-01-01	Zona Sur	Trucha	12.86	6.93	7.32	3.65	10.83	131.77
1	2015-01-02	Zona Norte	Pejerrey	14.62	7.12	7.06	11.63	2.45	132.82
2	2015-01-03	Zona Sur	Carachi	12.74	8.8	7.05			
3	2015-01-04	Zona Sur	Ispi	10.67	5.29	8.1			
4	2015-01-05	Zona Norte	Pejerrey	17.94	7.4	8.45			

Show 25 per page

Apéndice 5 Otros.

Operacionalización de las variables

Operacionalización de las variables

VARIABLE	DIMENSIÓN INSTRUMENTO	INDICADOR	
VARIABLE INDEPENDIENTE			
Algoritmo Modelo Predictivo	1.1. Condiciones Meteorológicas	1.1.1 Temperatura media del aire (°C).	Encuesta
		1.1.2 Precipitación acumulada diaria (mm).	
		1.1.3 Velocidad promedio y dirección predominante del viento (m/s, grados).	
	1.2. Condiciones Unaspológicas	1.2.1 Temperatura superficial del agua (°C).	
		1.2.2 Nivel de oxígeno disuelto (mg/L).	
	1.3. Contexto	1.2.3 Nivel de pH del agua.	
VARIABLE DEPENDIENTE	Temporal y Espacial	1.2.4 Transparencia del agua (profundidad de disco Secchi en cm).	
Eficiencia en la pesca artesanal	1.4. Esfuerzo y Método de Pesca	1.3.1 Período del año (temperatura: lluviosa/seca; mes).	
		1.3.2 Zona de pesca georreferenciada (coordenadas GPS o sector nominal: e.g., bahía Interior, zona abierta).	
		1.3.3 Profundidad promedio de la zona de faena (metros).	
	2.1. Rendimiento Productivo (Captura)	1.4.1 Número de embarcaciones por faena.	
		1.4.2 Horas netas de esfuerzo de pesca por faena.	
	2.2. Eficiencia Técnica (Productividad)	1.4.3 Tipo de arte de pesca utilizado (red de arrastre, red de cerco, línea).	
	2.3. Eficiencia Económica y Operativa	2.1.1 Volumen total de captura por faena (kg/faena).	
		2.1.2 Composición de la captura por especie principal (% de carachi, pejerrey, otras).	
		2.2.1 Captura por Unidad de Esfuerzo (CPUE): Kg capturados / embarcación.	
		2.2.2 Captura por Unidad de Esfuerzo (CPUE): Kg capturados / hora de pesca.	
		2.2.3 Tasa de éxito de faenas (% de faenas con captura > umbral económico).	
		2.3.1 Relación insumo-producto: Volumen de captura (kg) / litros de combustible consumidos.	
		2.3.2 Rendimiento económico	



captura por faena
(kg * precio
promedio de
venta).

2.1.1 Distancia
recorrida para
alcanzar la zona de
pesca (km).

Fuente: propia del autor

Condigo de la aplicación

1.

```
# Manipulación de
datos import
pandas as pd
import numpy as
np

# Visualización
import matplotlib.pyplot as
plt import seaborn as sns

# Preprocesamiento
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler,
OneHotEncoder from sklearn.compose import
ColumnTransformer
from sklearn.pipeline import Pipeline

# Modelos
from sklearn.linear_model import
LinearRegression from sklearn.ensemble import
RandomForestRegressor from
sklearn.neural_network import MLPRegressor

# Métricas
from sklearn.metrics import
mean_squared_error, mean_absolute_error,
r2_score
```



ANEXO 1 FORMULARIO DE AUTORIZACIÓN

AUTORIZACIÓN PARA LA INCORPORACIÓN DE LOS TRABAJOS DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL UANCV

Formato digital ☒

Fecha de entrega: 06/04/2026

1. Datos del autor (es):

Nombres y Apellidos: ALBERT LUIGHI FIGUEROA CARI

Dirección: Urb. Aziruni I Etapa Mz 12 Lt 23 Salcedo - Puno

DNI/Carné de Extranjería/Pasaporte N°: 70238459

Teléfono: 950710322 email: albertluighi@gmail.com

Nombres y Apellidos: _____

Dirección: _____

DNI/Carné de Extranjería/Pasaporte N°: _____

Teléfono: _____ email: _____

Facultad y/o Escuela de Posgrado: INGENIERÍA DE SISTEMAS

Escuela Profesional o Mención: INGENIERÍA DE SISTEMAS

Título o Grado Académico a optar: INGENIERO DE SISTEMAS

Asesor: Dr. PAUL MAMANI TISNADO

Esta obra se encuentra dentro de las siguientes denominaciones:

Trabajo de Investigación ☐ Tesis ☒ Trabajo de Suficiencia Profesional ☐ Trabajo Académico ☐

Título: APLICACIÓN DE ALGORITMOS PREDICTIVOS PARA MEJORAR LA EFICIENCIA EN LA PESCA ARTESANAL DE LA EMPRESA PUERTO LAGO PUNO 2024

Palabras claves, (3 a 5 términos): Machine learning, pesca artesanal, Lago Titicaca, modelo predictivo, eficiencia operativa, ciencia de datos, captura por unidad de esfuerzo (CPUE), Python, Random Forest, sostenibilidad pesquera

¿Esta obra se desarrolló en la UANCV ^{1, 2}?

2

¹ Indicar si su producción intelectual ha empleado recursos tales como, instalaciones, laboratorios, insumos, equipos, bases de datos, asesoría técnica por parte del personal de la UANCV, financiamiento, entre otros relacionados.

² Si su producción intelectual se desarrolló en la UANCV totalmente o parcialmente, deberá autorizar el depósito en el Repositorio de manera obligatoria.



2. Referencia de tesis:

☐ Bachiller ☒ Título ☐ 2da Especialidad ☐ Maestría ☐ Doctorado

3. Licencias:

a) Licencia estándar:

Bajo los siguientes términos, autorizo el depósito de mi tesis en el Repositorio Digital de la UANCV.

Con la autorización de depósito de mi producción Intelectual, otorgo a la Universidad Andina "Néstor Cáceres Velásquez" una licencia no exclusiva para reproducir, distribuir, comunicar al público, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público mi producción intelectual (incluido el resumen), en formato físico o digital, en cualquier medio, conocido o por conocerse, a través de los diversos servicios por la Universidad, creados o por crearse, tales como el Repositorio Digital de tesis UANCV, colección de producción intelectual, entre otros, en el Perú y en el extranjero por el tiempo y veces que considere necesarias, y libres de remuneraciones.

En virtud de dicha licencia, la Universidad Andina "Néstor Cáceres Velásquez" podrá reproducir mi producción intelectual en cualquier tipo de soporte y en más de un ejemplar, sin modificar su contenido, solo con propósitos de seguridad, respaldo y preservación.

Declaro que la producción intelectual es una creación de mi autoría y exclusiva titularidad, coautoría con titularidad compartida, y me encuentro facultado a conceder la presente licencia y, asimismo, garantizo que dicha producción intelectual no infringe derechos de autor de terceras personas.

La Universidad Andina "Néstor Cáceres Velásquez" consignará el nombre del y/o los autor(es) de la producción intelectual, y no le hará ninguna modificación más que la permitida en la licencia.

Autorizo su publicación (marque con una X)

- ☒ Sí, autorizo que se deposite inmediatamente.
- ☐ Sí, autorizo que se deposite a partir de la fecha (d/m/a): _____
- ☐ No autorizo.

b) Licencia CREATIVE COMMONS 4.0 INTERNACIONAL:

Si usted concede una licencia CREATIVE COMMONS sobre su producción intelectual, mantiene la titularidad de los derechos de autor de esta y, a la vez, permite que otras personas puedan reproducirla, comunicarla al público y distribuir ejemplares de esta, bajo las condiciones siguientes:

¿Quiere permitir usos comerciales de su producción intelectual?

Sí: significa que usted permite la reproducción, distribución y comunicación pública de la producción intelectual incluso con fines comerciales.

No: significa que usted permite la reproducción, y comunicación pública de la producción intelectual, pero sin fines comerciales.

- ☐ Sí autorizo
- ☒ No autorizo



Jurisdicción de su Licencia

Todas las licencias CREATIVE COMMONS son de ámbito mundial, sin embargo, usted puede elegir entre la opción "internacional" o una adaptada a su jurisdicción, como para el caso peruano.

La opción "internacional" emplea el lenguaje y la terminología de los tratados internacionales; en cambio, la adaptada a su jurisdicción, recoge las particularidades de la legislación peruana.

En consecuencia, **la opción "internacional" goza de una mayor eficacia a nivel mundial, gracias a que tiene jurisdicción neutral.** Mientras que la opción adaptada a la jurisdicción del Perú goza de una mayor eficacia ante los tribunales peruanos.

☐

Internacional

☒

Nacional

Línea de investigación: CIENCIAS DE LOS ORDENADORES – P24

Firma de Autor



huella digital

06 - ABRIL - 2026

Fecha